Bayesian Nonparametric System Reliability using Sets of Priors

Gero Walter[a], Louis J.M. Aslett[b], Frank P.A. Coolen[c]

[a]*School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, NL*
[b]*Department of Statistics, University of Oxford, Oxford, UK*
[c]*Department of Mathematical Sciences, Durham University, Durham, UK*

Abstract

An imprecise Bayesian nonparametric approach to system reliability with multiple types of components is developed. This allows modelling partial or imperfect prior knowledge on component failure distributions in a flexible way through bounds on the functioning probability. Given component level test data these bounds are propagated to bounds on the posterior predictive distribution for the functioning probability of a new system containing components exchangeable with those used in testing. The method further enables identification of prior-data conflict at the system level based on component level test data. New results on first-order stochastic dominance for the Beta-Binomial distribution make the technique computationally tractable. Our methodological contributions can be immediately used in applications by reliability practitioners as we provide easy to use software tools.

Keywords: System reliability, Survival signature, Imprecise probability, Bayesian nonparametrics, Prior-data conflict

1. Introduction

System reliability analysis is concerned with estimating the lifetime T_{sys} of complex systems. Usually, the goal is to determine the system reliability function $R_{\text{sys}}(t) = P(T_{\text{sys}} > t)$ based on the lifetime distributions of system components.

A critique of the methodological approach to a reliability analysis may often encompass a few common concerns. First, in a parametric setting, there

Email addresses: g.m.walter@tue.nl (Gero Walter), louis.aslett@stats.ox.ac.uk (Louis J.M. Aslett), frank.coolen@durham.ac.uk (Frank P.A. Coolen)

may be no particularly strong reason to believe that the small part of component model space covered by a particular probability distribution necessarily contains the 'true' component lifetime distribution. Further, Bayesian methods may be invoked in order to incorporate expert opinion or other knowledge which falls outside the specific testing data under consideration. The classic concern here is in whether one can truly express ones beliefs with the exactness a prior distribution requires. Finally, it would be valuable in application to have a means of identifying when the prior choice is having a strong effect and when not. Any method hoping to address these concerns must do so whilst enabling realistic system models (with heterogeneous component types) and remain computationally tractable.

Herein, we make steps toward addressing these concerns by developing a nonparametric method which utilises imprecise probability [4, 27] to model more vague or imperfect prior beliefs using upper and lower probabilities. This overcomes the concern about component lifetimes being outside a particular parametric family, uses a more flexible prior modelling framework and leads to an easy method of detecting conflicts between prior assumptions and observed failure times in test data. In the general context of Bayesian methods, this phenomenon is known as *prior-data conflict*, see, e.g., [17] or [9].

Furthermore, the method is based on the survival signature [11], a recent development which naturally accommodates heterogeneous component types laid out in any arbitrary manner. By separating the (time-invariant) system structure from the time-dependent failure probabilities of components, it allows straightforward and efficient computation of $R_{\text{sys}}(t)$.

Our imprecise probability approach provides bounds for $R_{\text{sys}}(t)$ by computing, for each t in an arbitrarily fine grid of time points $\mathcal{T}$, the posterior predictive probability interval for the event $T_{\text{sys}} > t$. Assuming the number of functioning components for each type and time t as binomially distributed, the intervals are derived from an imprecise Bayesian model using sets of conjugate Beta priors which allow to specify weak or partial prior information in an intuitive way. The width of the resulting posterior predictive probability intervals reflects the precision of the corresponding probability statements: a short range indicates that the system functioning probability can be quantified quite precisely, while a large range will indicate that our (probabilistic) knowledge is indeterminate. In particular, prior-data conflict leads to more cautious probability statements: When there is not enough data to overrule the prior, it is unclear whether to put more trust to prior assumptions or to the observations, and posterior inferences clearly reflect this state of uncertainty by larger ranges.

Our approach extends the literature on reliability with imprecise probability [for an overview see 26] by integrating the learning of component

reliabilities based on test data, in combination with the use of the survival signature. Approaches to system reliability based on generalizations of Bayesian networks like evidential networks [24] and credal networks [1] do usually not include test data in the model. Doing so is certainly possible in these frameworks, but this would add further to the complexity of the Bayesian network representation, in which updating (i.e., calculating posterior reliabilities) for each time t is generally NP-hard [20]. While determining the survival signature is also NP-hard, this has to be done only once in our approach. Furthermore, new results on first-order stochastic dominance for the Beta-Binomial distribution keep the need for numerical optimization in our model to a minimum.

In Section 2 we review the survival signature and in Section 3 we review the nonparametric approach to Bayesian reliability analysis upon which our work builds [3]. Section 4 details the reparametrisation of that approach which enables the natural formulation of the system reliability bounds, leading to nice closed form results in some later theory. Section 5 lays the ground work to incorporate imprecise probability, culminating in the main results and contributions of this work, detailed in Section 6. Section 7 provides details on the software contributions of this work and shows two worked examples demonstrating the practicality and usefulness of the method.

2. Survival Signature

In the mathematical theory of reliability, the main focus is on the functioning of a system given the functioning, or not, of its components and the structure of the system. The mathematical concept which is central to this theory is the *structure function* [5]. For a system with m components, let state vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \{0,1\}^m$, with $x_i = 1$ if the ith component functions and $x_i = 0$ if not. The labelling of the components is arbitrary but must be fixed to define $\boldsymbol{x}$. The structure function $\phi : \{0,1\}^m \to \{0,1\}$, defined for all possible $\boldsymbol{x}$, takes the value 1 if the system functions and 0 if the system does not function for state vector $\boldsymbol{x}$. Most practical systems are coherent, which means that $\phi(\boldsymbol{x})$ is non-decreasing in any of the components of $\boldsymbol{x}$, so system functioning cannot be improved by worse performance of one or more of its components. The assumption of coherent systems is also convenient from the perspective of uncertainty quantification for system reliability. It is further logical to assume that $\phi(\mathbf{0}) = 0$ and $\phi(\mathbf{1}) = 1$, so the system fails if all its components fail and it functions if all its components function.

For larger systems, working with the full structure function may be complicated, and one may particularly only need a summary of the structure

function in case the system has exchangeable components of one or more types. We use the term 'exchangeable components' to indicate that the failure times of the components in the system are exchangeable [16]. Coolen and Coolen-Maturi [11] introduced such a summary, called the *survival signature*, to facilitate reliability analyses for systems with multiple types of components. In case of just a single type of components, the survival signature is closely related to the system signature [22], which is well-established and the topic of many research papers during the last decade. However, generalization of the signature to systems with multiple types of components is extremely complicated (as it involves ordering order statistics of different distributions), so much so that it cannot be applied to most practical systems. In addition to the possible use for such systems, where the benefit only occurs if there are multiple components of the same types, the survival signature is arguably also easier to interpret than the signature.

Consider a system with $K \geq 1$ types of components, with m_k components of type $k \in \{1, \ldots, K\}$ and $\sum_{k=1}^{K} m_k = m$. Assume that the random failure times of components of the same type are exchangeable [16]. Due to the arbitrary ordering of the components in the state vector, components of the same type can be grouped together, leading to a state vector that can be written as $\boldsymbol{x} = (\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^K)$, with $\boldsymbol{x}^k = (x_1^k, x_2^k, \ldots, x_{m_k}^k)$ the sub-vector representing the states of the components of type k.

The *survival signature* for such a system, denoted by $\Phi(l_1, \ldots, l_K)$, with $l_k = 0, 1, \ldots, m_k$ for $k = 1, \ldots, K$, is defined as the probability for the event that the system functions given that *precisely* l_k of its m_k components of type k function, for each $k \in \{1, \ldots, K\}$ [11]. Essentially, this creates a K-dimensional partition for the event $T_{\text{sys}} > t$, such that $R_{\text{sys}}(t) = P(T_{\text{sys}} > t)$ can be calculated using the law of total probability:

$$\begin{aligned} P(T_{\text{sys}} > t) &= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} P(T_{\text{sys}} > t \mid C_t^1 = l_1, \ldots, C_t^K = l_K) \\ &\qquad \times P\Big(\bigcap_{k=1}^{K} \{C_t^k = l_k\} \Big) \\ &= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) P\Big(\bigcap_{k=1}^{K} \{C_t^k = l_k\} \Big), \end{aligned} \tag{1}$$

where $C_t^k \in \{0, 1, \ldots, m_k\}$ denotes the random number of components of type k functioning at time t.

For calculating the survival signature based on the structure function, observe that there are $\binom{m_k}{l_k}$ state vectors $\boldsymbol{x}^k$ with $\sum_{i=1}^{m_k} x_i^k = l_k$. Let $S_{l_k}^k$

denote the set of these state vectors for components of type k and let $S_{l_1,\ldots,l_K}$ denote the set of all state vectors for the whole system for which $\sum_{i=1}^{m_k} x_i^k = l_k$, $k = 1, \ldots, K$. Due to the exchangeability assumption for the failure times of the m_k components of type k, all the state vectors $\boldsymbol{x}^k \in S_{l_k}^k$ are equally likely to occur, hence [11]

$$\Phi(l_1, \ldots, l_K) = \left[\prod_{k=1}^{K} \binom{m_k}{l_k}^{-1} \right] \times \sum_{\boldsymbol{x} \in S_{l_1,\ldots,l_K}} \phi(\boldsymbol{x}) \,. \tag{2}$$

It should be emphasized that when using the survival signature, there are no restrictions on dependence of the failure times of components of different types, as the probability $P(\bigcap_{k=1}^{K} \{C_t^k = l_k\})$ can take any form of dependence into account, for example one can include common-cause failures quite straightforwardly into this approach [12]. However, there is a substantial simplification if one can assume that the failure times of components of different types are independent, and even more so if one can assume that the failure times of components of type k are conditionally independent and identically distributed with CDF $F_k(t)$. With these assumptions, we get

$$R_{\mathrm{sys}}(t) = \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \left[\Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} \left(\binom{m_k}{l_k} [F_k(t)]^{m_k - l_k} [1 - F_k(t)]^{l_k} \right) \right] .$$

We will employ both assumptions in this paper, leading to C_t^k having a Beta-Binomial distribution, giving us a closed form expression for $P(C_t^k = l_k)$ for all t, k, and l_k.

The main advantage of the survival signature, in line with this property of the signature for systems with a single type of components [22], is that the information about the system structure is fully separated from the information about functioning of the components, which simplifies related statistical inference as well as considerations of optimal system design. In particular for study of system reliability over time, with the structure of the system, and hence the survival signature, not changing, this separation also enables relatively straightforward statistical inferences.

There are several relatively straightforward generalizations of the use of the survival signature. The probabilities for the numbers of functioning components can be generalized to lower and upper probabilities, as e.g. done by Coolen et al. [14] within the nonparametric predictive inference (NPI) framework of statistics [10], where lower and upper probabilities for the events $C_k = l_k$ are inferred from test data on components of the same types as those in the system. This is an approach that is also followed in the current

paper, but with the use of generalized Bayesian inference instead of NPI. Like Coolen et al. [14], we will utilize the monotonicity of the survival signature for coherent systems to simplify computations.

While there are, as mentioned above, no restrictions on dependence of failure times of components of different types that can be reflected by the survival signature method, some possible dependencies will require thoughtful consideration. For example, it may be the case that components that are near to each other in the system have a stronger dependency of their failure times than components that are further apart. If this is the case there are several options, the two main ones are as follows. First, if there is detailed knowledge about such dependence (as e.g. modelled in popular spatial statistics approaches) then there are probably no components that can be grouped, hence one has to use the full structure function. In practice, this may be feasible for small systems, but for large systems one would probably still wish to make some exchangeability assumptions in order to enable analysis, where it is understood that this is only approximate. The second case results from this, namely that one can still define groups of components which have exchangeable failure times and which are related in the same way to other groups of components to facilitate the dependence modelling. For example, one might consider components in a large electricity network to have differing failure time characteristics depending on physical location, e.g. windturbines near hills or at sea, while apart from the location aspect one would consider their failure times exchangeable. In such cases, these are represented by different groups in the survival signature approach.

3. Nonparametric Bayesian Approach for Component Reliability

Let us denote the random failure time of component number i of type k by T_i^k, $i = 1, \ldots, m_k$. The failure time distribution can be written in terms of the cdf $F^k(t) = P(T_i^k \leq t)$, or in terms of the reliability function $R^k(t) = P(T_i^k > t) = 1 - F^k(t)$, also known as the survival function. For a nonparametric description of $R^k(t)$, we consider a set of time points t, $t \in \mathcal{T} = \{t_1, \ldots, t_{\max}\}$.

At each time point t, the operational state of a single component of type k is Bernoulli distributed (functioning: 1, failed: 0) with parameter p_t^k, so that

$$P\big(\mathbb{I}(T_i^k > t) = 1\big) = p_t^k\,,$$
$$P\big(\mathbb{I}(T_i^k > t) = 0\big) = 1 - p_t^k\,,$$

That is, $\mathbb{I}(T_i^k > t) \sim \text{Bernoulli}(p_t^k)$, $i = 1, \ldots, m_k$, $t \in \mathcal{T}$.

The set of probabilities $\{p_t^k, t \in \mathcal{T}\}$ defines a discrete failure time distribution for components of type k through

$$R^k(t_j) = P(T^k > t_j) = p_{t_j}^k,\ t_j = t_1, \ldots, t_{\max}\,.$$

We can also express this failure time distribution through the probability mass function (pmf) and discrete hazard function,

$$\begin{aligned} f^k(t_j) &= P\big(T^k \in (t_j, t_{j+1}]\big) = p_{t_j}^k - p_{t_{j+1}}^k\,, \\ h^k(t_j) &= P\big(T^k \in (t_j, t_{j+1}] \mid T^k > t_j\big) = \frac{f^k(t_j)}{R^k(t_j)}\,. \end{aligned}$$

The time grid $\mathcal{T}$ can be chosen to be appropriately dense for the application at hand, where a simple extension between grid points can be made by taking $R^k(\cdot)$ to be the right continuous step function induced by the grid values, $R^k(t) = p_{t_j}^k, t \in [t_j, t_{j+1})$, or by taking $p_{t_j}^k$ and $p_{t_{j+1}}^k$ as upper and lower bounds for $R^k(t)$, $t \in [t_j, t_{j+1})$.

The independence assumption for components of the same type immediately implies that the number of functioning components of type k in the system is binomially distributed, $C_t^k = \sum_{i=1}^{m_k} \mathbb{I}(T_i^k > t) \sim \text{Binomial}(p_t^k, m_k)$.

The sequence of p_t^k's can, in theory, be directly chosen to arbitrarily closely approximate any valid lifetime pdf on $[0, \infty)$, for example matching a bathtub curve for the corresponding hazard rate $h^k(t_j)$. Naturally, $p_{t_j}^k \geq p_{t_{j+1}}^k$ should hold (assuming no repair). However, such direct specification is non-trivial, neglects any inherent uncertainty in the particular choice, and cannot be easily combined with test data. To account for the uncertainty, one can express knowledge about p_t^k through a prior distribution. A convenient and natural choice is $p_t^k \sim \text{Beta}(\alpha_t^k, \beta_t^k)$, particularly because in a Bayesian inferential setting this is the conjugate prior which leads to a Beta posterior.

Let the lifetime test data collected on component k be $\boldsymbol{t}^k = (t_1^k, \ldots, t_{n_k}^k)$. At each fixed time $t \in \mathcal{T}$, this corresponds to an observation from the Binomial model described above, $s_t^k = \sum_{i=1}^{n_k} \mathbb{I}(t_i^k > t)$. The posterior is then $p_t^k \mid s_t^k \sim \text{Beta}(\alpha_t^k + s_t^k, \beta_t^k + n_k - s_t^k)$. The combination of a Binomial observation model with a Beta prior is often called Beta-Binomial model.

The posterior predictive distribution for the number of components surviving at time t in a new system, based on the lifetime data and the prior information, is then given by a so-called Beta-Binomial distribution, $C_t^k \mid s_t^k \sim \text{Beta-Binomial}(m_k, \alpha_t^k + s_t^k, \beta_t^k + n_k - s_t^k)$. That is, we have

$$P(C_t^k = l_k \mid s_t^k) = \binom{m_k}{l_k} \frac{B(l_k + \alpha_t^k + s_t^k, m_k - l_k + \beta_t^k + n_k - s_t^k)}{B(\alpha_t^k + s_t^k, \beta_t^k + n_k - s_t^k)}\,,$$

where $B(\cdot, \cdot)$ is the Beta function. This posterior predictive distribution is essentially a Binomial distribution where the functioning probability p_t^k takes values in $[0, 1]$ with the probability given by the posterior on p_t^k.

Consequently, in order to calculate the system reliability according to (1), for each component type k we need lifetime data $\boldsymbol{t}^k$, and have to choose $2 \times |\mathcal{T}|$ parameters to specify the prior distribution for the discrete survival function of T_i^k. In total, values for $2 \times |\mathcal{T}| \times K$ parameters must be chosen.

4. Reparametrisation of the Beta Distribution

The parametrisation of the Beta distribution used above is common, and allows α_t^k and β_t^k to be interpreted as hypothetical numbers of functioning and failed components of type k at time t, respectively. However, as recognized by [27, §5.3], when we generalise to sets of priors in the sequel, it is useful to consider a different parametrisation.

For clarity of presentation we will temporarily drop the super- and subscript k and t indices for component type and time. Instead of α and β, we consider the parameters $n^{(0)} \in [0, \infty)$ and $y^{(0)} \in [0, 1]$, where

$$n^{(0)} = \alpha + \beta \qquad \text{and} \qquad y^{(0)} = \frac{\alpha}{\alpha + \beta}\,, \tag{3}$$

or equivalently, $\alpha = n^{(0)} y^{(0)}$ and $\beta = n^{(0)}(1 - y^{(0)})$. The upper index $^{(0)}$ is used to identify these as prior parameter values, in contrast to their posterior values $n^{(n)}$ and $y^{(n)}$ obtained after observing n failure times (see below). $n^{(0)}$ and $y^{(0)}$ are sometimes called *canonical* parameters, identified from rewriting the density in canonical form; see for example [8, pp. 202 and 272f], or [29, §1.2.3.1]. This canonical form gives a common structure to all conjugacy results in exponential families.

From the properties of the Beta distribution, it follows that $y^{(0)} = \mathrm{E}[p]$ is the prior expectation for the functioning probability p, and that larger $n^{(0)}$ values lead to greater concentration of probability measure around $y^{(0)}$, since $\mathrm{Var}(p) = \frac{y^{(0)}(1-y^{(0)})}{n^{(0)}+1}$. Consequently, $n^{(0)}$ represents the prior strength and moreover can be directly interpreted as a pseudocount, as will become clear. Indeed, consider the posterior given that s out of n components function: by conjugacy $p \mid s$ is Beta distributed with updated parameters

$$n^{(n)} = n^{(0)} + n\,, \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{s}{n}\,. \tag{4}$$

Thus, after observing that s out of n components function (at time t), the posterior mean $y^{(n)}$ for p is a weighted average of the prior mean $y^{(0)}$ and

s/n (the fraction of functioning components in the data), with weights proportional to $n^{(0)}$ and n, respectively. Therefore $n^{(0)}$ takes on the same role for the prior mean $y^{(0)}$ as the sample size n does for the observed mean s/n, leading to the notion of it being a pseudocount.

Reintroducing time and component type indices, the posterior predictive Beta-Binomial probability mass function (pmf) can be written in terms of the updated parameters as

$$P(C_t^k = l_k \mid s_t^k) = \binom{m_k}{l_k} \frac{B(l_k + n_{k,t}^{(n)} y_{k,t}^{(n)}, m_k - l_k + n_{k,t}^{(n)}(1 - y_{k,t}^{(n)}))}{B(n_{k,t}^{(n)} y_{k,t}^{(n)}, n_{k,t}^{(n)}(1 - y_{k,t}^{(n)}))}\,, \tag{5}$$

with the corresponding cumulative mass function (cmf) given by

$$F_{C_t^k \mid s_t^k}(l_k) = P(C_t^k \leq l_k \mid s_t^k) = \sum_{j_k=0}^{l_k} P(C_t^k = j_k \mid s_t^k)\,. \tag{6}$$

The parameterisation in terms of prior mean and prior strength (or pseudocount) makes clear that in this conjugate setting, learning from data corresponds to averaging between prior and data. This form is attractive not only because it enhances the interpretability of the model and prior specification, but crucially it also makes clear what should be a serious concern in any Bayesian analysis: when observed data differ greatly from what is expressed in the prior, this conflict is simply averaged out and is not reflected in the posterior or posterior predictive distributions.

As a simple example, imagine that we expect p_t^k to be about 0.75 for a certain k and t, so we choose $y_{k,t}^{(0)} = 0.75$, and that we value this choice of mean functioning probability with $n_{k,t}^{(0)} = 8$, i.e., equivalently to having seen 8 observations with a mean 0.75. If we observe $n_k = 16$ components of type k in the test data and $s_t^k = 12$ function at time t, then $s_t^k/n_k = 0.75$ as we expect, so that the updated parameters are $n_{k,t}^{(n)} = 24, y_{k,t}^{(n)} = 0.75$. However, in contrast, unexpectedly observing that no component functions at time t instead leads to parameters $n_{k,t}^{(n)} = 24, y_{k,t}^{(n)} = 0.25$. The prior and the posteriors based on these two scenarios are depicted in the left panels of Figure 1, along with their corresponding predictive Beta-binomial pmf and cmf for the case $m_k = 5$ (right panels).

Due to symmetry, we see that both posteriors have the same variance, although arising from two fundamentally different scenarios. Posterior 1 is based on data exactly according to prior expectations; the increase in confidence on $p_t^k \approx 0.75$ is reflected in a more concentrated posterior density, and the posterior predictive is changed only slightly. However, it may be

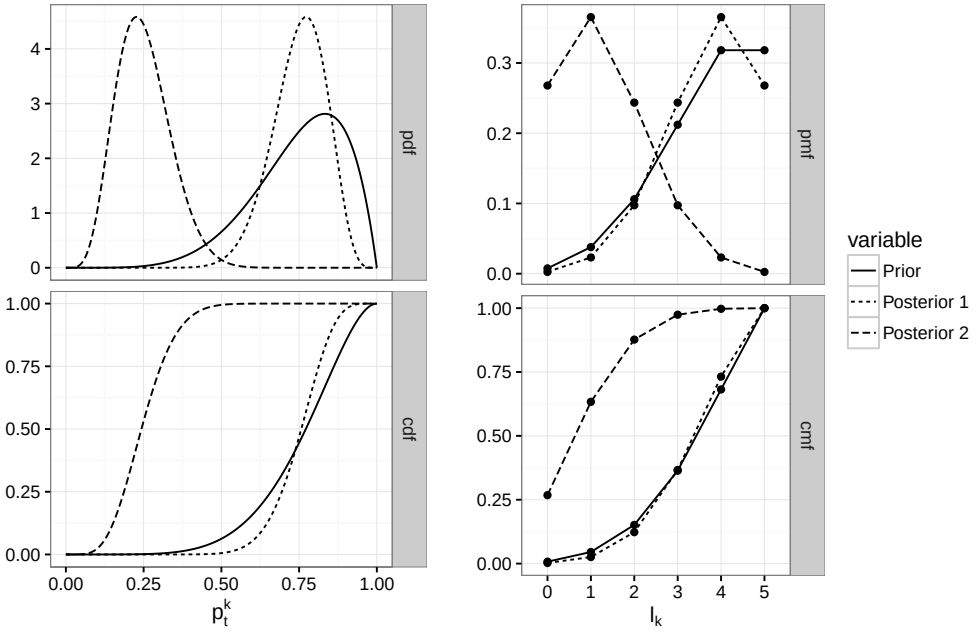

Figure 1: Beta densities (top left) and cdfs (bottom left), with the corresponding Beta-binomial predictive probability mass functions (top right) and cumulative mass functions (bottom right), for a prior with $n_{k,t}^{(0)} = 8, y_{k,t}^{(0)} = 0.75$, and posteriors based on $n_t^k = 16$ observations with $s_t^k = 12$ (Posterior 1) and $s_t^k = 0$ (Posterior 2), respectively. Data for Posterior 1 confirm prior assumptions, while data for Posterior 2 are in conflict with the prior. However, this conflict is averaged out, and Posterior 1 and Posterior 2 have the same spread, both in the posterior pdf/cdf and the posterior predictive pmf/cmf, such that Posterior 2 gives a false sense of certainty despite the massive conflict between prior and data.

cause for concern to see the same degree of confidence in Posterior 2, which is based on data that is in sharp conflict with prior expectations. Posterior 2 places most probability weight around 0.25, averaging between prior expectation and data, with the same variance as Posterior 1. Accordingly, rather than conveying the conflict between observed and expected functioning probabilities with increased variance, Posterior 2 instead gives a false sense of certainty.

To enable diagnosis of when this undesirable behaviour occurs, we propose to use an imprecise probability approach based on sets of Beta priors, described in the following section.

5. Sets of Beta Priors

As was shown by Walter and Augustin [30], we can have both tractability and meaningful reaction to prior-data conflict by using sets of priors $\mathcal{M}_{k,t}^{(0)}$ produced by parameter sets $\Pi_{k,t}^{(0)} = [\underline{n}_{k,t}^{(0)}, \overline{n}_{k,t}^{(0)}] \times [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$ (a detailed discussion of different choices for $\Pi_{k,t}^{(0)}$ is given in Walter [29, §3.1].) In our model, each prior parameter pair $(n_{k,t}^{(0)}, y_{k,t}^{(0)}) \in \Pi_{k,t}^{(0)}$ corresponds to a Beta prior, thus $\mathcal{M}_{k,t}^{(0)}$ is a set of Beta priors. The set of posteriors $\mathcal{M}_{k,t}^{(n)}$ is obtained by updating each prior in $\mathcal{M}_{k,t}^{(0)}$ according to Bayes' Rule. This element-by-element updating can be rigorously justified as ensuring coherence [27, §2.5], and was termed "Generalized Bayes' Rule" by Walley [27, §6.4]. Due to conjugacy, $\mathcal{M}_{k,t}^{(n)}$ is a set of Beta distributions with parameters $(n_{k,t}^{(n)}, y_{k,t}^{(n)})$, obtained by updating $(n_{k,t}^{(0)}, y_{k,t}^{(0)}) \in \Pi_{k,t}^{(0)}$ according to (4), leading to the set of updated parameters

$$\Pi_{k,t}^{(n)} = \left\{ (n_{k,t}^{(n)}, y_{k,t}^{(n)}) \mid (n_{k,t}^{(0)}, y_{k,t}^{(0)}) \in \Pi_{k,t}^{(0)} = [\underline{n}_{k,t}^{(0)}, \overline{n}_{k,t}^{(0)}] \times [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}] \right\}. \tag{7}$$

Examples for parameter sets $\Pi_{k,t}^{(0)}$ and $\Pi_{k,t}^{(n)}$ as in (7) are depicted in Figure 2. Such rectangular prior parameter sets $\Pi_{k,t}^{(0)}$ have been shown to balance desirable model properties and ease of elicitation (see Walter [29, pp. 123f] or Troffaes et al. [25]). For each component type k and time point t, one need only specify the four parameters $\underline{n}_{k,t}^{(0)}, \overline{n}_{k,t}^{(0)}, \underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}$ (so in total $4 \times |\mathcal{T}|$ parameters are needed to define the set of prior distributions on the survival function of each component).

A desirable inference property arising from this setup is that the posterior parameter set $\Pi_{k,t}^{(n)}$ is not rectangular in the way that the prior parameter set is. Indeed, the shape of $\Pi_{k,t}^{(n)}$ depends on the presence or absence of prior-data conflict, which is naturally operationalised as $s_t^k / n_k \notin [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$: that is, prior-data conflict is defined to occur when, at time t, the observed fraction of functioning components is outside its *a priori* expected range.

First, in the absence of prior-data conflict, $\Pi_{k,t}^{(n)}$ shrinks in the $y_{k,t}$ dimension; how much it shrinks depending on $n_{k,t}^{(0)} \in [\underline{n}_{k,t}^{(0)}, \overline{n}_{k,t}^{(0)}]$, leading to the so-called spotlight shape depicted in Figure 2 (left). Since $y_{k,t}^{(n)}$ gives the posterior expectation for the functioning probability p_t^k, shorter $y_{k,t}^{(n)}$ intervals mean more precise knowledge about p_t^k. Also, the variance interval for p_t^k (not shown) will shorten and shift towards zero, as the Beta distributions in $\mathcal{M}_{k,t}^{(n)}$ will be more concentrated due to the increase of $n_{k,t}^{(0)}$ to $n_{k,t}^{(n)}$.

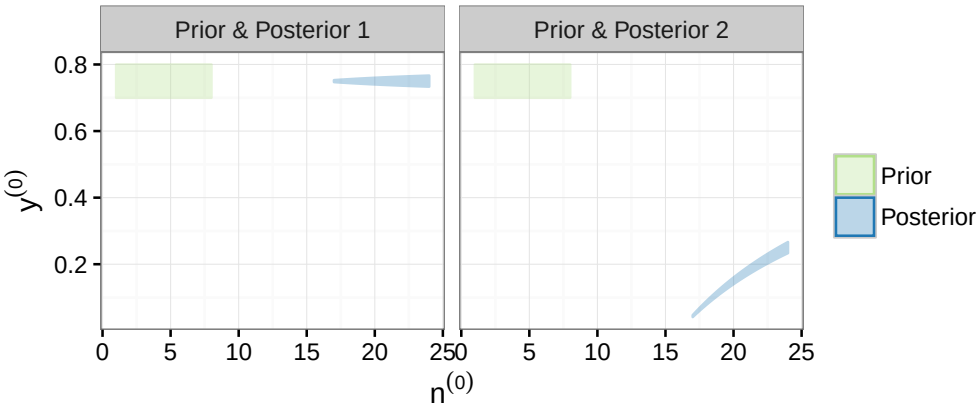

Figure 2: Prior parameter set $\Pi_{k,t}^{(0)} = [1, 8] \times [0.7, 0.8]$ and posterior parameter set $\Pi_{k,t}^{(n)}$ for data $s_t^k/n_k = 12/16$ (Posterior 1, left) and $s_t^k/n_k = 0/16$ (Posterior 2, right). For no-conflict data ($s_t^k/n_k \in [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$), $\Pi_{k,t}^{(n)}$ has the 'spotlight' shape (left); in case of prior-data conflict ($s_t^k/n_k \notin [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$), $\Pi_{k,t}^{(n)}$ has the 'banana' shape (right), leading to a large degree of imprecision in the $y_{k,t}^{(n)}$ dimension of $\Pi_{k,t}^{(n)}$, thus reflecting increased uncertainty about the functioning probability p_t^k due to the conflict between prior assumptions and observed data.

Alternatively, when there is conflict between prior and observed data (i.e. $s_t^k/n_k \notin [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$), $\Pi_{k,t}^{(n)}$ instead adopts the so-called 'banana shape', arising from the intervals for $y_{k,t}^{(n)}$ being shifted closer to s_t^k/n_k for lower $n_{k,t}^{(n)}$ values than for higher $n_{k,t}^{(n)}$ values, see Figure 2 (right). Overall, this results in a wider $y_{k,t}^{(n)}$ interval compared to the no conflict case, reflecting the extra uncertainty due to prior-data conflict. In other words, the posterior sets make more cautious probability statements about p_t^k, as desired in this scenario.

Based on these shapes and (4), it is possible to deduce the following expressions for the lower and upper bounds of $y_{k,t}^{(n)}$:

$$\begin{aligned} \min_{\Pi_{k,t}^{(n)}} y_{k,t}^{(n)} &= \begin{cases} \left(\overline{n}_{k,t}^{(0)} \underline{y}_{k,t}^{(0)} + s_t^k\right) / \left(\overline{n}_{k,t}^{(0)} + n_k\right) & \text{if } s_t^k/n_k \geq \underline{y}_{k,t}^{(0)} \\ \left(\underline{n}_{k,t}^{(0)} \underline{y}_{k,t}^{(0)} + s_t^k\right) / \left(\underline{n}_{k,t}^{(0)} + n_k\right) & \text{if } s_t^k/n_k < \underline{y}_{k,t}^{(0)} \end{cases}, \\ \max_{\Pi_{k,t}^{(n)}} y_{k,t}^{(n)} &= \begin{cases} \left(\overline{n}_{k,t}^{(0)} \overline{y}_{k,t}^{(0)} + s_t^k\right) / \left(\overline{n}_{k,t}^{(0)} + n_k\right) & \text{if } s_t^k/n_k \leq \overline{y}_{k,t}^{(0)} \\ \left(\underline{n}_{k,t}^{(0)} \overline{y}_{k,t}^{(0)} + s_t^k\right) / \left(\underline{n}_{k,t}^{(0)} + n_k\right) & \text{if } s_t^k/n_k > \overline{y}_{k,t}^{(0)} \end{cases}. \end{aligned} \tag{8}$$

Note that the lower bound for $y_{k,t}^{(n)}$ is always attained at $\underline{y}_{k,t}^{(0)}$, the upper bound at $\overline{y}_{k,t}^{(0)}$. Also note that when $s_t^k/n_k \in [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$, both the lower and the upper bounds for $y_{k,t}^{(n)}$ are attained at $\overline{n}_{k,t}^{(0)}$, corresponding to the spotlight

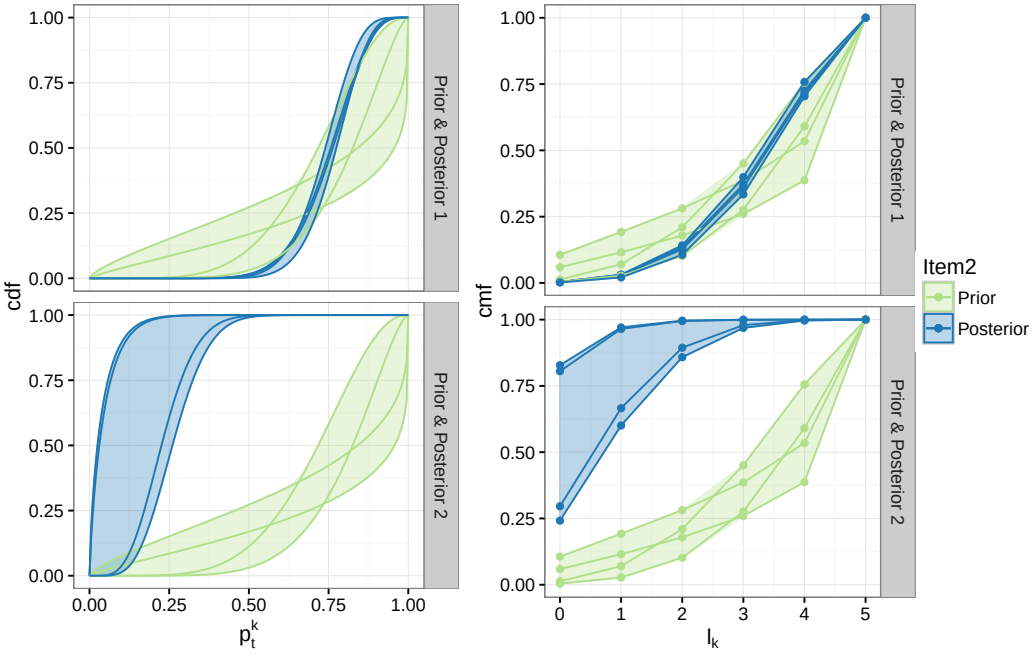

Figure 3: Sets of Beta pdfs (left) and Beta-Binomial cmfs (right, for $m_k = 5$) corresponding to the prior and posterior parameter sets in Figure 2. The sets are depicted as shaded areas, with the distributions corresponding to the four corners of the prior parameter set $\Pi_{k,t}^{(0)}$ (or their posterior counterparts) as solid lines. The top row depicts the set of prior cdfs/cmfs and the set of posterior cdfs/cmfs for the case where data confirm prior assumptions (see left panel of Figure 2); the bottom row depicts the (identical) set of prior cdfs/cmfs and the set of posterior cdfs/cmfs in case of prior-data conflict (see right panel of Figure 2). The set of posterior cdfs and cmfs is much larger in case of prior-data conflict: uncertainty due to this conflict is reflected through increased imprecision.

shape. However, when $s_t^k/n_k \notin [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$, the banana shape indicates that one of the bounds for $y_{k,t}^{(n)}$ is attained at $\underline{n}_{k,t}^{(0)}$.

The different locations and sizes of $\Pi_{k,t}^{(n)}$ in the conflict versus no conflict case are then, in turn, also reflected in the corresponding sets of Beta cdfs and Beta-Binomial cmfs. As an example, those corresponding to the parameter sets in Figure 2 are depicted in Figure 3.

In the no conflict case (Posterior 1, top row), the reduction of the $y_{k,t}^{(n)}$ range in $\Pi_{k,t}^{(n)}$ leads to a much smaller set of Beta and Beta-Binomial distributions. For example, the range of predictive probabilities that two out of a set of five components of type k function at time t has changed from $[0.10, 0.28]$ *a priori* to $[0.11, 0.14]$ *a posteriori*. This reflects the gain in precision due to test data in accordance with prior assumptions.

In contrast, for the prior-data conflict case (Posterior 2, bottom row),

the wide $y_{k,t}^{(n)}$ range in $\Pi_{k,t}^{(n)}$ leads to a set of Beta and Beta-Binomial distributions that is much larger than in the no conflict case. Here, the range of posterior predictive probabilities that two out of a set of five components of type k function at time t is now $[0.86, 1.00]$ *a posteriori*, i.e., less precise than in the no conflict case. Using sets of Beta priors, the resulting set of posterior predictive Beta-Binomial distributions reflects the precision of prior information, the amount of data, and prior-data conflict.

Furthermore, with sets of Beta priors it is also possible to express prior ignorance by letting $\underline{y}_{k,t}^{(0)} \to 0$ and $\overline{y}_{k,t}^{(0)} \to 1$ for some or all $t \in \mathcal{T}$. Such models are called *near-noninformative* or *near-ignorance models* in the imprecise probability literature (see Walley [27, §5.3.2] for Beta priors, [28] for Dirichlet priors, and Benavoli and Zaffalon [6, 7] for exponential family priors), as they provide vacuous bounds only for a certain class of inferences, and the choice for the prior strength parameter $n_{k,t}^{(0)}$ influences posterior inferences. (Note that it is not advisable to choose $\underline{y}_{k,t}^{(0)} = 0$ and $\overline{y}_{k,t}^{(0)} = 1$, as this can lead to improper posterior predictive distributions. For example, at any $t < \min(\boldsymbol{t}^k)$, we would have $\overline{y}_{k,t}^{(n)} = 1$, leading to one argument of the Beta function in the denominator of (5) being zero.) The limits $\underline{y}_{k,t}^{(0)} \to 0$ and $\overline{y}_{k,t}^{(0)} \to 1$ imply we are only prepared to give trivial bounds for the functioning probability and do not wish to commit to any specific knowledge about p_t^k *a priori*. This provides a more natural choice of 'noninformative' prior over $[0, 1]$ than the usual choice of a Beta prior with $\alpha_t^k = \beta_t^k = 1$ (or $n_{k,t}^{(0)} = 2$, $y_{k,t}^{(0)} = 0.5$). Such a prior for all $t \in \mathcal{T}$ actually reflects a belief that the component reliability function is on average $1/2$ for all t, which is not an expression of ignorance, but rather a very specific (and arguably peculiar) prior belief.

In a near-noninformative setting, the choice of $\underline{n}_{k,t}^{(0)}$ is not relevant, because (8) implies both lower and upper bound for $y_{k,t}^{(n)}$ are obtained with $\overline{n}_{k,t}^{(0)}$. In particular, $\underline{y}_{k,t}^{(0)} > 0$ and $\overline{y}_{k,t}^{(0)} < 1$ can be chosen such that $\frac{s_t^k}{n_k} \in [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$ for all $t \in \big(\min(\boldsymbol{t}^k), \max(\boldsymbol{t}^k)\big)$. Naturally, one cannot have prior-data conflict in cases of near prior ignorance.

6. Sets of System Reliability Functions

The elements reviewed and extended above culminate hereinafter in the primary contribution of the current work, providing a framework in which the nonparametric Bayesian system reliability approach developed in [3] is extended to sets of system reliability functions by incorporating the sets of priors approach of Walter and Augustin [30]. This allows for partial or vague

specification of prior component reliability functions, and enables diagnosis of prior-data conflict which is consequential at the system level.

6.1. Computation of bounds

To obtain the lower and upper bound for the system reliability function $R_{\text{sys}}(t)$, we now need to minimise and maximise Equation (1) over $\Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}$ for each t, where the posterior predictive probabilities for C_t^k are given by the Beta-Binomial pmf (5). We therefore have

$$
\begin{aligned}
&\underline{R}_{\text{sys}}(t \mid \boldsymbol{t}^1, \ldots, \boldsymbol{t}^K) \\
&= \min_{\Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}} R_{\text{sys}}(t \mid \Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}, \boldsymbol{t}^1, \ldots, \boldsymbol{t}^K) \\
&= \min_{\Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}} \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid y_{k,t}^{(0)}, n_{k,t}^{(0)}, s_t^k) \\
&= \min_{\Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}} \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \times \\
&\qquad \prod_{k=1}^{K} \binom{m_k}{l_k} \frac{B(l_k + n_{k,t}^{(n)} y_{k,t}^{(n)}, m_k - l_k + n_{k,t}^{(n)}(1 - y_{k,t}^{(n)}))}{B(n_{k,t}^{(n)} y_{k,t}^{(n)}, n_{k,t}^{(n)}(1 - y_{k,t}^{(n)}))} \\
&= \min_{\Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}} \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \times \\
&\qquad \prod_{k=1}^{K} \binom{m_k}{l_k} \frac{B(l_k + n_{k,t}^{(0)} y_{k,t}^{(0)} + s_t^k, m_k - l_k + n_{k,t}^{(0)}(1 - y_{k,t}^{(0)}) + n_k - s_t^k)}{B(n_{k,t}^{(0)} y_{k,t}^{(0)} + s_t^k, n_{k,t}^{(0)}(1 - y_{k,t}^{(0)}) + n_k - s_t^k)},
\end{aligned}
\tag{9}
$$

and similarly maximising for $\overline{R}_{\text{sys}}(\cdot)$. In doing so, we assume components of the same type k to be exchangeable given $\Pi_{k,t}^{(0)}$ [see 4, §3.4 on extending exchangeability to imprecise probability], and components of different types to follow *strong independence* [see 4, §3.2.4].

Note that $\Phi(\cdot)$ is non-decreasing in each of its arguments $l_1, \ldots, l_K$, thus if there is first-order stochastic ordering on $P(C_t^k = l_k \mid y_{k,t}^{(0)}, n_{k,t}^{(0)}, s_t^k)$ for each k, then this ordering can be used to determine the elements of $\Pi_{k,t}^{(0)}$ which minimise and maximise the overall system reliability function without resorting to computationally expensive exhaustive searches or numerical optimisation.

We therefore start by providing the following result, where indices are suppressed for readability. We use $\geq_{\text{st}}$ to denote first-order stochastic dominance.

Theorem 1. *Let β_y denote the Beta-Binomial distribution with probability mass function parameterised as:*

$$p(l \mid y, n, m, s, N) \propto \frac{B(l + ny + s, m - l + n(1 - y) + N - s)}{B(ny + s, n(1 - y) + N - s)},$$

with $n, m, s,$ and N fixed and unknown.

Then $\beta_{\overline{y}} \geq_{\text{st}} \beta_{\underline{y}}$ $\forall$ $\overline{y} > \underline{y}$ with $\overline{y}, \underline{y} \in (0, 1)$.

The proof is provided in Appendix A, p.30.

Consequently, for each component, the posterior predictive Beta Binomial distributions with larger prior functioning probability stochastically dominate those with smaller prior functioning probability, providing rigorous proof which accords with intuition. Applying this result to the sets of system reliability functions, together with the monotonicity in the survival signature, means that $\underline{R}_{\text{sys}}(\cdot)$ is attained when $y_{k,t}^{(0)} = \underline{y}_{k,t}^{(0)}$ and $\overline{R}_{\text{sys}}(\cdot)$ is attained when $y_{k,t}^{(0)} = \overline{y}_{k,t}^{(0)}$ for all possible $n_{k,t}^{(0)}$ values.

The analogous result for $n_{k,t}^{(0)}$ is more subtle, because stochastic dominance is not guaranteed at a single value. The following Theorem provides simple sufficient conditions under which an upper or lower limit has first-order stochastic dominance and has virtually no computational overhead to test.

Theorem 2. *Let β_n denote the Beta-Binomial distribution with probability mass function parameterised as:*

$$p(l \mid y, n, m, s, N) \propto \frac{B(l + ny + s, m - l + n(1 - y) + N - s)}{B(ny + s, n(1 - y) + N - s)},$$

with $y, m, s,$ and N fixed and unknown. Then,

$$y > \frac{s + m - 1}{N + m - 1} \implies \beta_{\overline{n}} \geq_{\text{st}} \beta_{\underline{n}}$$

and

$$y < \frac{s}{N + m - 1} \implies \beta_{\overline{n}} \leq_{\text{st}} \beta_{\underline{n}}$$

The proof is provided in Appendix A, p.31.

If $\frac{s}{N+m-1} < y < \frac{s+m-1}{N+m-1}$ then Theorem 2 cannot determine stochastic dominance. The following Lemma which is slightly more computationally costly, but still much faster than an exhaustive search, may be able to determine first-order stochastic dominance in such situations.

Lemma 3. *Let β_n denote the Beta-Binomial distribution as in Theorem 2. Define*

$$\mathcal{L}_{\overline{n},\underline{n}}(l) := \frac{p(l \mid y, \overline{n}, m, s, N)}{p(l \mid y, \underline{n}, m, s, N)}$$

Then,

$$\left.\begin{array}{r} \mathcal{L}_{\overline{n},\underline{n}}(0) \leq 1 \\ \mathcal{L}_{\overline{n},\underline{n}}(m) \geq 1 \end{array}\right\} \implies \beta_{\overline{n}} \geq_{\text{st}} \beta_{\underline{n}}$$

and

$$\left.\begin{array}{r} \mathcal{L}_{\overline{n},\underline{n}}(0) \geq 1 \\ \mathcal{L}_{\overline{n},\underline{n}}(m) \leq 1 \end{array}\right\} \implies \beta_{\overline{n}} \leq_{\text{st}} \beta_{\underline{n}}$$

The proof is provided in Appendix A, p.32. In the cases where neither Theorem 2 or Lemma 3 apply, the entire posterior system reliability function must be optimised to find the minima/maxima. In practice, in the examples to be presented in the sequel, Theorem 2 and Lemma 3 do provide guarantees of first-order stochastic dominance for the vast majority of time points, t, substantially lowering the computational costs of performing the minimisation/maximisation involved in finding the sets of system reliability functions compared to either numerical optimisation or an exhaustive grid search (which would get exponentially slower in the number of different components). Appendix B provides a detailed discussion of exactly how often these bounds hold and how often one must resort to numerical optimisation.

Thus, due to Theorems 1, 2 and Lemma 3, (9) transforms to

$$\begin{aligned}
&\underline{R}_{\text{sys}}(t \mid \boldsymbol{t}^1, \ldots, \boldsymbol{t}^K) \\
&= \min_{\Pi_{1,t}^{(0)}, \ldots, \Pi_{K,t}^{(0)}} \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \times \\
&\qquad \prod_{k=1}^{K} \binom{m_k}{l_k} \frac{B(l_k + n_{k,t}^{(0)} y_{k,t}^{(0)} + s_t^k, m_k - l_k + n_{k,t}^{(0)}(1 - y_{k,t}^{(0)}) + n_k - s_t^k)}{B(n_{k,t}^{(0)} y_{k,t}^{(0)} + s_t^k, n_{k,t}^{(0)}(1 - y_{k,t}^{(0)}) + n_k - s_t^k)} \\
&= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \times \\
&\qquad \prod_{k=1}^{K} \binom{m_k}{l_k} \frac{B(l_k + \widetilde{n}_{k,t}^{(0)} \underline{y}_{k,t}^{(0)} + s_t^k, m_k - l_k + \widetilde{n}_{k,t}^{(0)}(1 - \underline{y}_{k,t}^{(0)}) + n_k - s_t^k)}{B(\widetilde{n}_{k,t}^{(0)} \underline{y}_{k,t}^{(0)} + s_t^k, \widetilde{n}_{k,t}^{(0)}(1 - \underline{y}_{k,t}^{(0)}) + n_k - s_t^k)}
\end{aligned} \tag{10}$$

where

$$\widetilde{n}_{k,t}^{(0)} = \begin{cases} \overline{n}_{k,t}^{(0)} & \text{if } \underline{y}_{k,t}^{(0)} < \dfrac{s_t^k}{n_k + m_k - 1} \vee \left(\mathcal{L}_{\overline{n}_{k,t}^{(0)}, \underline{n}_{k,t}^{(0)}}(0) \geq 1 \wedge \mathcal{L}_{\overline{n}_{k,t}^{(0)}, \underline{n}_{k,t}^{(0)}}(m) \leq 1 \right) \\ \underline{n}_{k,t}^{(0)} & \text{if } \underline{y}_{k,t}^{(0)} > \dfrac{s_t^k + m_k - 1}{n_k + m_k - 1} \vee \left(\mathcal{L}_{\overline{n}_{k,t}^{(0)}, \underline{n}_{k,t}^{(0)}}(0) \leq 1 \wedge \mathcal{L}_{\overline{n}_{k,t}^{(0)}, \underline{n}_{k,t}^{(0)}}(m) \geq 1 \right) \\ & \text{optimised otherwise} \end{cases}$$

The result for $\overline{R}_{\text{sys}}(\cdot)$ is completely analogous. It is interesting to note that if $m_k = 1$ the bounds are sharp on stochastic dominance. In particular, when $m_k = 1$, $\underline{y}_{k,t}^{(0)} < \frac{s_t^k}{n_k}$ indicates the lower bound is not in conflict with the observed data, whilst $\underline{y}_{k,t}^{(0)} > \frac{s_t^k}{n_k}$ is in conflict since the observed empirical probability of functioning at time t is below the prior lower bound. Consequently, note that $\underline{n}_{k,t}^{(0)}$ is used only when the prior comes into conflict with the data. Since $n_{k,t}^{(0)}$ controls the prior certainty, this accords with the intuition that the least certain prior bound is invoked when in a conflict setting and the more certain prior bound used when the data agrees.

6.2. Prior parameter choice

In the following, we will give some guidelines on how to choose the parameter sets $\Pi_{k,1}^{(0)}, \ldots, \Pi_{k,t_{\max}}^{(0)}$ which define the set of prior discrete reliability functions for components of type k. We advocate that this is much easier in terms of $n^{(0)}$ and $y^{(0)}$ than it would be in terms of α and β.

As mentioned in Section 3, the functioning probabilities p_t^k must satisfy $p_{t_j}^k \geq p_{t_{j+1}}^k$.This naturally translates to conditions on the prior for p_t^k, so that for example $\overline{y}_{k,t_j}^{(0)} \geq \overline{y}_{k,t_{j+1}}^{(0)}$ and $\underline{y}_{k,t_j}^{(0)} \geq \underline{y}_{k,t_{j+1}}^{(0)}$ should hold. Because s_t^k/n_k is decreasing in t, the weighted average property of the update step in Equation (4) for $y_{k,t}$ ensures that $\overline{y}_{k,t_j}^{(n)} \geq \overline{y}_{k,t_{j+1}}^{(n)}$ and $\underline{y}_{k,t_j}^{(n)} \geq \underline{y}_{k,t_{j+1}}^{(n)}$. In situations where one has a high degree of certainty about the functioning probability for low t, but less certainty about what happens for larger t, then one can let $\underline{y}_{k,t}^{(0)}$ drop to (almost) 0, but clearly $\overline{y}_{k,t}^{(0)}$ should not increase.

It is inadvisable to express certainty in the expected functioning probabilities with $n_{k,t}^{(0)}$ bounds that vary substantially over the range of t. With (strongly) differing $n_{k,t}^{(0)}$ bounds, monotonicity of the $y_{k,t}^{(n)}$ bounds cannot be guaranteed. For example, if $\overline{y}_{k,t_j}^{(0)} = \overline{y}_{k,t_{j+1}}^{(0)}$, $\underline{y}_{k,t_j}^{(0)} = \underline{y}_{k,t_{j+1}}^{(0)}$, and $s_{t_j}^k/n_k \in [\underline{y}_{k,t_j}^{(0)}, \overline{y}_{k,t_j}^{(0)}]$ (meaning there is no prior-data conflict), then should there be no observed failures in $[t_j, t_{j+1}]$, so that $s_{t_{j+1}}^k/n_k = s_{t_j}^k/n_k$, then

$$\overline{n}_{k,t_j}^{(0)} < \overline{n}_{k,t_{j+1}}^{(0)} \quad \Longrightarrow \quad \overline{y}_{k,t_j}^{(n)} < \overline{y}_{k,t_{j+1}}^{(n)} \quad \text{and}$$

$$\overline{n}^{(0)}_{k,t_j} > \overline{n}^{(0)}_{k,t_{j+1}} \quad \Longrightarrow \quad \underline{y}^{(n)}_{k,t_j} < \underline{y}^{(n)}_{k,t_{j+1}}$$

Again, this follows from (4), the weighted average property. It is possible to construct similar examples with regard to the lower bound $\underline{n}^{(0)}_{k,t_j}$. Therefore, we advise taking the same $n^{(0)}_{k,t}$ bounds for all t as far as possible. If they do change, it must be very gradual and we recommend diagnosing any problems as above.

Generally, the interpretation as pseudocount or prior strength should guide the choice of bounds for $n^{(0)}_{k,t}$; low values for $n^{(0)}_{k,t}$ as compared to the test sample size n_k give low weight to the prior expected functioning probability intervals $[\underline{y}^{(0)}_{k,t}, \overline{y}^{(0)}_{k,t}]$, and the location of posterior intervals $[\underline{y}^{(n)}_{k,t}, \overline{y}^{(n)}_{k,t}]$ will be dominated by the location of s^k_t/n_k. Furthermore, the length of $[\underline{y}^{(n)}_{k,t}, \overline{y}^{(n)}_{k,t}]$ is shorter for low $n^{(0)}_{k,t}$ values. Specifically, in a no-conflict situation, when $\overline{n}^{(0)}_{k,t} = n_k$ then $[\underline{y}^{(n)}_{k,t}, \overline{y}^{(n)}_{k,t}]$ has half the length of $[\underline{y}^{(0)}_{k,t}, \overline{y}^{(0)}_{k,t}]$. In contrast, high values for $n^{(0)}_{k,t}$ will lead to slower learning and wider $y^{(n)}_{k,t}$ intervals, which means more cautious posterior inferences. The difference betweeen $\overline{n}^{(0)}_{k,t}$ and $\underline{n}^{(0)}_{k,t}$ determines the strength of the prior-data conflict sensitivity; as is clear from Figure 2 and (8), the wider the $n^{(0)}_{k,t}$ interval, the wider $[\underline{y}^{(n)}_{k,t}, \overline{y}^{(n)}_{k,t}]$ in case of conflict. So it seems useful to choose $\underline{n}^{(0)}_{k,t} = 1$ or $\underline{n}^{(0)}_{k,t} = 2$, while choosing $\overline{n}^{(0)}_{k,t}$ with help of the half-width rule as described above.

As mentioned in Section 5, it is not advisable to choose $\underline{y}^{(0)}_{k,t} = 0$ and $\overline{y}^{(0)}_{k,t} = 1$. For any $t \notin \big(\min(\boldsymbol{t}^k), \max(\boldsymbol{t}^k)\big)$, this can lead to improper posterior predictive distributions. However, it is possible to choose values close to 0 and 1, respectively, and due to the linear update step (4) for $y^{(n)}_{k,t}$, posterior inferences are not overly sensitive to whether $\overline{y}^{(n)}_{k,t} = 0.99$ or $\overline{y}^{(n)}_{k,t} = 0.9999$. Likewise, our nonparametric method does not cause unintuitive tail behaviour as some parametric methods do; there is no problem, for example, with assigning $\overline{y}^{(n)}_{k,t}$ near-zero for large t if prior knowledge suggests so.

While it is possible to set the bounds $\underline{y}^{(0)}_{k,t}$ and $\overline{y}^{(0)}_{k,t}$ for each $t \in \mathcal{T}$ individually, in practice this will be often too time-consuming when $\mathcal{T}$ forms a dense grid. Switching to a coarser time grid will waste information from data, as then failure times in the test data are rounded up to the next $t \in \mathcal{T}$. In the examples here we elicit bounds for a subset of $\mathcal{T}$ and fill up the time grid with the least committal bounds, i.e., taking $\overline{y}^{(0)}_{k,t}$ equal to last (in the time sequence) elicited $\overline{y}^{(0)}_{k,t}$, and likewise $\underline{y}^{(0)}_{k,t}$ equal to next (in time sequence) elicited $\underline{y}^{(0)}_{k,t}$. A possible elicitation procedure in this vein could be to start

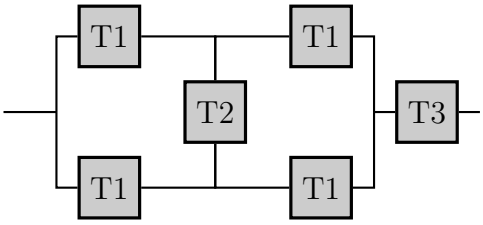

Figure 4: Reliability block diagram for a 'bridge' system with three component types.

with eliciting $y_{k,t}^{(0)}$ bounds for a few 'central' time points t, filling up the grid as described above accordingly, and then to further refine the obtained bounds as deemed necessary by the expert.

7. Practical Usage and Examples

7.1. Software

The methods of this paper have been implemented in the R [21] package `ReliabilityTheory` [2], providing an easy to use interface for reliability practitioners. The primary function, which computes the upper and lower posterior predictive system survival probabilities as in (10), is named `nonParBayesSystemInferencePriorSets()`. The user specifies the times at which to evaluate the bounds, the survival signature ($\Phi(\cdot)$), the component test data ($\boldsymbol{t}^1, \ldots, \boldsymbol{t}^K$), and the prior parameter set for each component type and time ($\Pi_{k,t}^{(0)}$, via $\overline{n}_{k,t}^{(0)}, \underline{n}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}$, and $\underline{y}_{k,t}^{(0)}$). All computations of $\underline{R}_{\text{sys}}$ and $\overline{R}_{\text{sys}}$ at different time points are performed in parallel automatically when the CPU has multiple cores and making automatic use of the theoretical results in Section 6 where applicable, performing exhaustive search in the few cases they are not.

Note that computation of the system signature itself can be simplified by expressing the structure of the system as an undirected graph using the `computeSystemSurvivalSignature()` function in the same package, leaving only data and prior to be handled. These publicly available functions have been used in computing all the following examples for reproducibility. See Appendix C for further details of how to use this software.

7.2. Examples

7.2.1. Toy example

As a toy example, consider a 'bridge' type system layout with three types of components T1, T2 and T3, as depicted in Figure 4. The survival signature for this system is given in Table 1. All rows with T3 = 0 have been omitted; without T3, the system cannot function, thus $\Phi = 0$. For

T1	T2	T3	Φ	T1	T2	T3	Φ
0	0	1	0	0	1	1	0
1	0	1	0	1	1	1	0
2	0	1	0.33	2	1	1	0.67
3	0	1	1	3	1	1	1
4	0	1	1	4	1	1	1

Table 1: Survival signature for the bridge system from Figure 4, omitting all rows with T3 = 0, since $\Phi = 0$ for these.

t	$[0,1)$	$[1,2)$	$[2,3)$	$[3,4)$	$[4,5)$
$\underline{y}_{3,t}^{(0)}$	0.625	0.375	0.250	0.125	0.010
$\overline{y}_{3,t}^{(0)}$	0.999	0.875	0.500	0.375	0.250

Table 2: Lower and upper prior functioning probability bounds for component type T3 in the 'bridge' system example.

component types T1 and T2, we consider a near-noninformative set of prior reliability functions. For components of type T3, we consider an informative set of prior reliability functions as given in Table 2. This set could result from eliciting prior functioning probabilities at times $0, 1, 2, 3, 4, 5$ only, and filling up the rest. These prior assumptions, together with sets of posterior reliability functions resulting from three different scenarios for test data for component type T3, are illustrated in Figures 5, 6 and 7; test data for components of type T1 and T2 are invariably taken as $\boldsymbol{t}^1 = (2.2, 2.4, 2.6, 2.8)$ and $\boldsymbol{t}^2 = (3.2, 3.4, 3.6, 3.8)$, respectively.

In Figure 5, test data for component type T3 is $\boldsymbol{t}^3 = (0.5, 1.5, 2.5, 3.5)$, and so in line with expectations. The posterior set of reliability functions for each component type and the whole system is considerably smaller compared to the prior set (due to the low prior strength intervals $[\underline{n}_{1,t}^{(0)}, \overline{n}_{1,t}^{(0)}] = [\underline{n}_{2,t}^{(0)}, \overline{n}_{2,t}^{(0)}] = [1, 2]$, $[\underline{n}_{3,t}^{(0)}, \overline{n}_{3,t}^{(0)}] = [1, 4]$) and so giving more precise reliability statements. We see that posterior lower and upper functioning probabilities drop at those times t when there is a failure time in the test data, or a drop in the prior functioning probability bounds. Note that the lower bound for the prior system reliability function is zero due to the prior lower bound of zero for T1; for the system to function, at least two components of type T1 must function.

In Figure 6, test data of component type T3 is $\boldsymbol{t}^3 = (0.6, 0.7, 0.8, 0.9)$, and so earlier than expected. Compared to Figure 5, posterior functioning intervals for T3 are wider between $t = 1$ and $t = 3.5$, reflecting additional

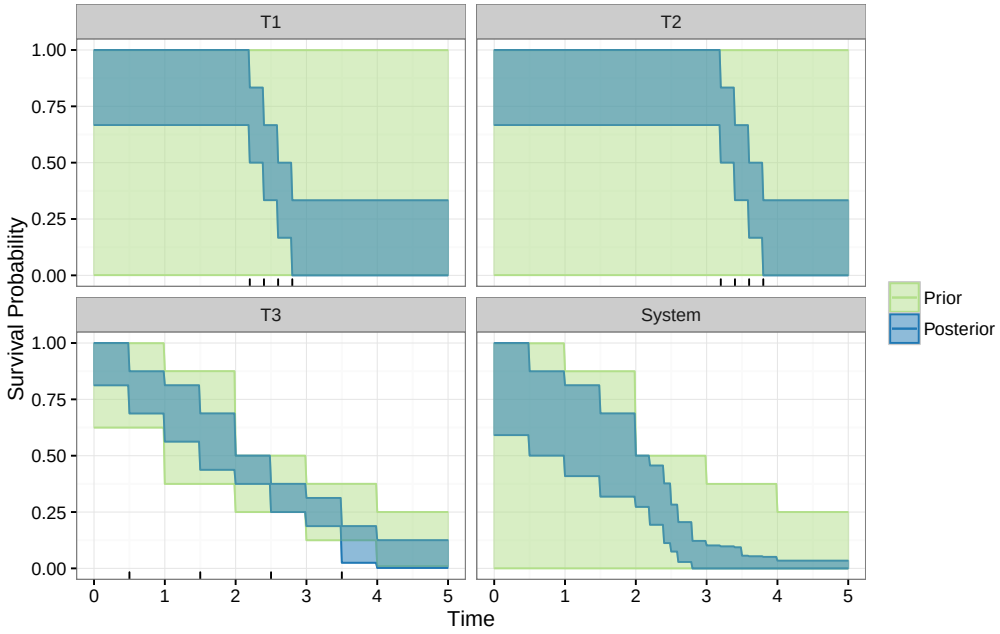

Figure 5: Prior and posterior sets of reliability functions for the 'bridge' system and its three component types, with failure times as expected for component type T3. Test data failure times are denoted with tick marks near the time axis.

imprecision due to prior-data conflict. For $t > 1$, it is clearly visible how $\overline{y}_{3,t}^{(n)}$ is halfway between $\overline{y}_{3,t}^{(0)}$ and $s_t^3/n_3 = 0$ (weights $\overline{n}_{3,t}^{(0)} = 4$ and $n_3 = 4$), while $\underline{y}_{3,t}^{(n)}$ is one-fifth of $\underline{y}_{3,t}^{(0)}$ (weights $\underline{n}_{3,t}^{(0)} = 1$ and $n_3 = 4$). Note that the posterior system functioning probability is constant for $t \in [1, 2]$ because in that interval the prior functioning probability is constant and there are no observed failures.

In Figure 7, test data of component type T3 is $\boldsymbol{t}^3 = (4.1, 4.2, 4.3, 4.4)$, and so observed failures are later than expected. Here we see that for $t \in [2, 4]$, posterior functioning bounds for T3 are even wider than prior functioning bounds. The width turns back to being half the prior width only after the four failures. The imprecision carries over to the system bounds, where we see wider bounds as compared to the other two scenarios especially between $t = 2$ and $t = 4$. In particular, also note that at the system level posterior bounds are a subset of prior bounds after $t = 2.6$, although prior-data conflict for the individual component type T3 extends well beyond $t = 4$. This demonstrates the power of this technique to identify prior-data conflict which is actually consequential at the system level, not just the component level — in other words, for mission times $t > 2.6$, we can diagnose that the prior-data conflict need not be of elevated concern for this system viewed as a whole.

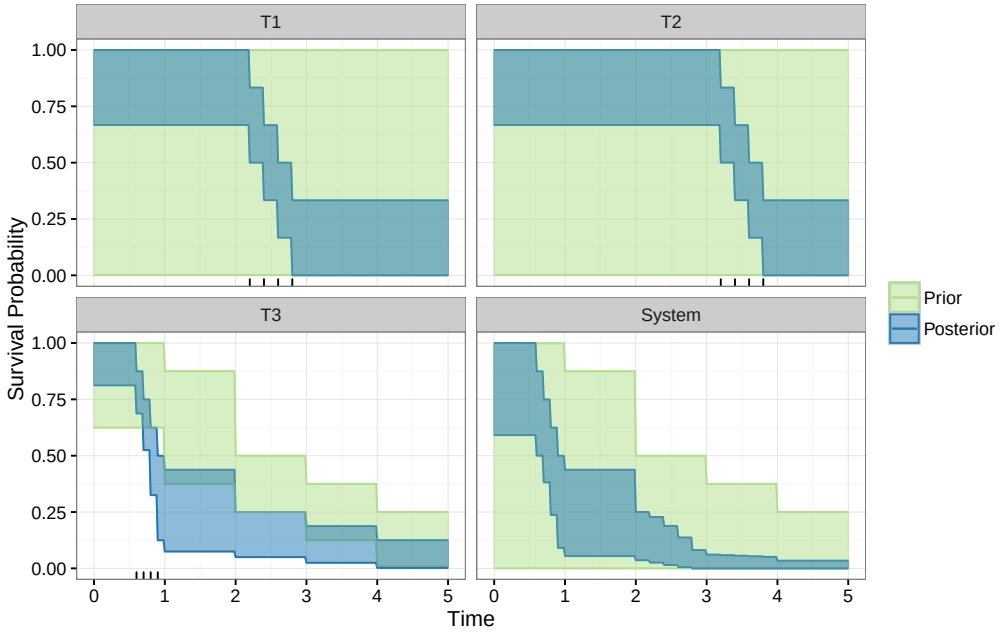

Figure 6: Prior and posterior sets of reliability functions for the 'bridge' type system and its three component types, with failure times earlier as expected for component type T3.

Nevertheless, the posterior system reliability bounds are wider than in the no-conflict case for $t \in [1, 4.4]$, signalising the general need for caution in this scenario.

7.2.2. Automotive brake system

We also consider a simplified automotive brake system. The master brake cylinder (M) activates all four wheel brake cylinders (C1 – C4), which in turn actuate a braking pad assembly each (P1 – P4). The hand brake mechanism (H) goes directly to the brake pad assemblies P3 and P4; the car brakes when at least one brake pad assembly is actuated. All values for $\Phi \notin \{0, 1\}$ are given in Table 3. The system layout is depicted in Figure 8, together with prior and posterior sets of reliability functions for the four component types and the complete system. Observed lifetimes from test data are indicated by tick marks in each of the four component type panels, where $n_{\mathrm{M}} = 5$, $n_{\mathrm{H}} = 10$, $n_{\mathrm{C}} = 15$, and $n_{\mathrm{P}} = 20$. We consider 301 evenly spaced time points t on $[0, 10]$, assume $[\underline{n}^{(0)}_{\mathrm{M},t}, \overline{n}^{(0)}_{\mathrm{M},t}] = [1, 8]$ $\forall t$, and $[\underline{n}^{(0)}_{k,t}, \overline{n}^{(0)}_{k,t}] = [1, 2]$ for $k \in \{\mathrm{H}, \mathrm{C}, \mathrm{P}\}$ and all t. Prior functioning probability bounds for M are based on a Weibull cdf with shape 2.5 and scales 6 and 8 for the lower and upper bound, respectively. The prior bounds for P can be seen as the least committal bounds derived from an expert statement of $y^{(0)}_{\mathrm{P},t} \in [0.5, 0.65]$ for

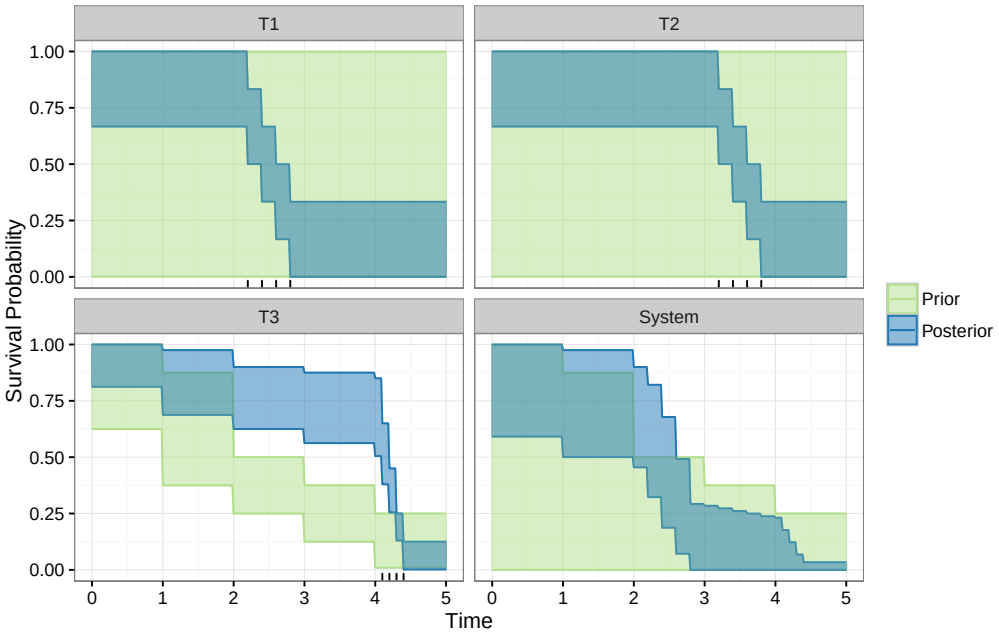

Figure 7: Prior and posterior sets of reliability functions for the 'bridge' type system and its three component types, with failure times later as expected for component type T3.

M	H	C	P	Φ	M	H	C	P	Φ
1	0	1	1	0.25	1	0	2	1	0.50
1	0	1	2	0.50	1	0	2	2	0.83
1	0	1	3	0.75	1	0	3	1	0.75
0	1	0	1	0.50	1	1	0	1	0.50
0	1	0	2	0.83	1	1	0	2	0.83
0	1	1	1	0.62	1	1	1	1	0.62
0	1	1	2	0.92	1	1	1	2	0.92
0	1	2	1	0.75	1	1	2	1	0.75
0	1	2	2	0.97	1	1	2	2	0.97
0	1	3	1	0.88	1	1	3	1	0.88

Table 3: Survival signature values $\notin \{0, 1\}$ for the simplified automotive brake system depicted in Figure 8.

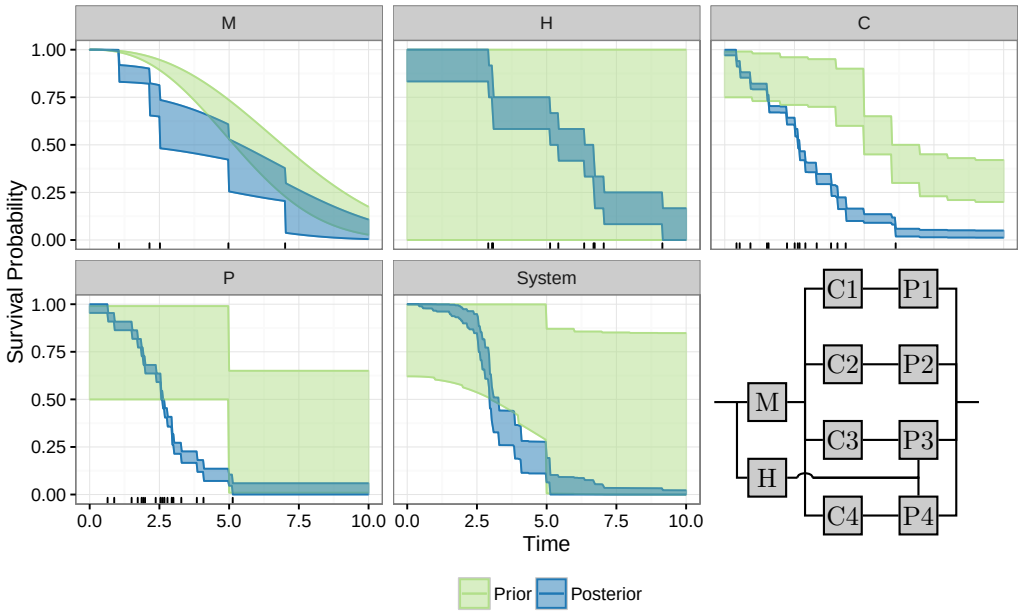

Figure 8: Prior and posterior sets of reliability functions for a simplified automotive brake system with layout as depicted in the lower right panel.

$t = 5$ only. For H, near-noninformative prior functioning probability bounds have been selected; with the upper bound for P being approximately one for $t \leq 5$ as well, the prior upper system reliability bound for $t \leq 5$ is close to one, too, since the system can function on H and one of P3 and P4 alone. Note that the posterior functioning probability interval for M is wide not only due to the limited number of observations, but also because $\overline{n}^{(0)}_{\mathrm{M},t} = 8$ and the prior-data conflict reaction.

Posterior functioning probability bounds for the complete system are much more precise than the prior system bounds, reflecting the information gained from component test data. The posterior system bounds can be also seen to reflect location and precision of the component bounds; for example, the system bounds drop drastically between $t = 2.5$ and $t = 3.5$ mainly due to the drop of the bounds for P at that time.

It is also interesting to note that the prior-data conflict which is consequential at the system level occurs over roughly the same range in t as there is prior-data conflict for component type P. Indeed, this occurs despite there being prior-data conflict in both M and C over much larger ranges, giving valuable insight into which prior requires further expert attention — thus the technique avoids wasted time addressing prior-data conflict in components

which may not be relevant when propagated to the uncertainty in the whole system.

In this example, the theory presented in Section 6 enabled avoiding numerical optimisation in 91.9% of cases for $\underline{R}_{\text{sys}}(\cdot)$ and in 95.7% of cases for $\overline{R}_{\text{sys}}(\cdot)$. Further detail is in Appendix B.3.

8. Conclusions

In this paper we have contributed an imprecise Bayesian nonparametric approach to system reliability with multiple types of components. The approach allows modelling partial or imperfect prior knowledge on component failure distributions in a flexible way through bounds on the functioning probability for a grid of time points, and combines this information with test data in an imprecise Bayesian framework. Component-wise predictions on the number of functioning components are then combined to bounds for the system survival probability by means of the survival signature. New results on first-order stochastic dominance for the Beta-Binomial distribution enable closed-form solutions for these bounds in most cases and avoid exponential growth in the complexity of computing the estimate as the number of components grows. The widths of the resulting system reliability bounds reflect the amount of test data, the precision of prior knowledge, and crucially provide an easily used method to identify whether these two information sources are in conflict in a way which is of consequence to the whole system reliability estimate.

These methodological contributions can be immediately used in applications by reliability practitioners as we provide easy to use software tools.

An important next step is to extend the model to include right-censored observations which are common in the reliability setting. In particular, this allows to use component failure observations from a running system to calculate its remaining useful life. We see two potential approaches. First, to obtain lower and upper system reliability bounds one can assume that a component either fails immediately after censoring or continues to function during the entire time horizon. This minimal assumption will be simple to implement but will lead to high imprecision. Alternatively, one can assume exchangeability with other surviving components at the moment of censoring. This approach will be more complex to accommodate but will lead to less imprecision. Indeed, this assumption lies at the core of the Kaplan-Meier estimator [18], and has already been adopted by Coolen and Yan [15] in an imprecise probability context.

Upscaling the survival signature to large real-world systems and networks, consisting of thousands of components, is a major challenge. However, even

for such systems the fact that one only needs to derive the survival signature once for a system is an advantage, and also the monotonicity of the survival signature for coherent systems is very useful if one can only derive it partially.

The survival signature and its use for uncertainty quantification for system reliability can be generalized quite straightforwardly, mainly due to the simplicity of this concept. For example, one may generalize the system structure function from a binary function to a probability [see 13], to reflect uncertainty about system functioning for known states of its components, with a further generalization to imprecise probabilities possible.

Acknowledgements

The authors are grateful for support of their work from two anonymous reviewers and for their comments which led to improved presentation.

Gero Walter was supported by the DINALOG project CAMPI ("Coordinated Advanced Maintenance and Logistics Planning for the Process Industries").

Louis Aslett was supported by the i-like project (EPSRC grant reference number EP/K014463/1).

References

[1] Antonucci, A., 2011. The imprecise noisy-OR gate. In: FUSION '11: Proceedings of the 14th International Conference on Information Fusion. IEEE, pp. 709–715.

[2] Aslett, L., 2016. ReliabilityTheory: Tools for structural reliability analysis. R package.
URL `http://www.louisaslett.com`

[3] Aslett, L., Coolen, F., Wilson, S., 2015. Bayesian inference for reliability of systems and networks using the survival signature. Risk Analysis 35, 1640–1651.
URL `http://dx.doi.org/10.1111/risa.12228`

[4] Augustin, T., Coolen, F., de Cooman, G., Troffaes, M., 2014. Introduction to Imprecise Probabilities. Wiley, New York.

[5] Barlow, R., Proschan, F., 1975. Statistical Theory of Reliability and Life Testing. Holt, Rinehart and Winston, Inc., New York.

[6] Benavoli, A., Zaffalon, M., 2012. A model of prior ignorance for inferences in the one-parameter exponential family. Journal of Statistical Planning and Inference 142, 1960–1979.
URL http://dx.doi.org/10.1016/j.jspi.2012.01.023

[7] Benavoli, A., Zaffalon, M., 2015. Prior near ignorance for inferences in the k-parameter exponential family. Statistics 49 (5), 1104–1140.
URL http://dx.doi.org/10.1080/02331888.2014.960869

[8] Bernardo, J., Smith, A., 2000. Bayesian Theory. Wiley, Chichester.

[9] Bickel, D., 2015. Inference after checking multiple Bayesian models for data conflict and applications to mitigating the influence of rejected priors. International Journal of Approximate Reasoning 66, 53–72.
URL http://dx.doi.org/10.1016/j.ijar.2015.07.012

[10] Coolen, F., 2011. Nonparametric predictive inference. In: Lovric, M. (Ed.), International Encyclopedia of Statistical Science. Springer, Berlin, pp. 968–970.

[11] Coolen, F., Coolen-Maturi, T., 2012. Generalizing the signature to systems with multiple types of components. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (Eds.), Complex Systems and Dependability. Vol. 170 of Advances in Intelligent and Soft Computing. Springer, pp. 115–130.

[12] Coolen, F., Coolen-Maturi, T., 2015. Predictive inference for system reliability after common-cause component failures. Reliability Engineering and System Safety 135, 27–33.

[13] Coolen, F., Coolen-Maturi, T., 2016. The structure function for system reliability as predictive (imprecise) probability. Reliability Engineering and System Safety, to appear.

[14] Coolen, F., Coolen-Maturi, T., Al-nefaiee, A., 2014. Nonparametric predictive inference for system reliability using the survival signature. Journal of Risk and Reliability 228, 437–448.

[15] Coolen, F., Yan, K., 2004. Nonparametric predictive inference with right-censored data. Journal of Statistical Planning and Inference 126, 25–54.

[16] De Finetti, B., 1974. Theory of Probability. Wiley, Chichester.

[17] Evans, M., Moshonov, H., 2006. Checking for prior-data conflict. Bayesian Analysis 1, 893–914.
URL http://projecteuclid.org/euclid.ba/1340370946

[18] Kaplan, E., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457–481.

[19] Klenke, A., Mattner, L., 2010. Stochastic ordering of classical discrete distributions. Advances in Applied Probability 42 (2), 392–410.

[20] Mauá, D., de Campos, C., Benavoli, A., Antonucci, A., 2014. Probabilistic inference in credal networks: new complexity results. Journal of Artificial Intelligence Research 50, 603–637.

[21] R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

[22] Samaniego, F., 2007. System Signatures and their Applications in Engineering Reliability. Springer, New York.

[23] Shaked, M., Shanthikumar, J., 2007. Stochastic orders, 1st Edition. Springer, New York.

[24] Simon, C., Weber, P., Evsukoff, A., 2008. Bayesian networks inference algorithm to implement Dempster Shafer theory in reliability analysis. Reliability Engineering & System Safety 93, 950–963.
URL http://dx.doi.org/10.1016/j.ress.2007.03.012

[25] Troffaes, M., Walter, G., Kelly, D., 2013. A robust Bayesian approach to modelling epistemic uncertainty in common-cause failure models. Reliability Engineering & System Safety 125, 13–21.
URL http://dx.doi.org/10.1016/j.ress.2013.05.022

[26] Utkin, L., Coolen, F., 2007. Imprecise reliability: an introductory overview. In: Levitin, G. (Ed.), Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability. Springer, Berlin, pp. 261–306.

[27] Walley, P., 1991. Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, London.

[28] Walley, P., 1996. Inferences from multinomial data: Learning about a bag of marbles. Journal of the Royal Statistical Society, Series B 58, 3–34.

[29] Walter, G., 2013. Generalized Bayesian inference under prior-data conflict. Ph.D. thesis, Department of Statistics, LMU Munich.
URL http://edoc.ub.uni-muenchen.de/17059/

[30] Walter, G., Augustin, T., 2009. Imprecision and prior-data conflict in generalized Bayesian inference. Journal of Statistical Theory and Practice 3, 255–271.
URL http://dx.doi.org/10.1080/15598608.2009.10411924

Appendix

A. Proofs

Proof of Theorem 1, p16. Consider the likelihood ratio for the two Beta Binomial distributions $\beta_{\overline{y}}$ and $\beta_{\underline{y}}$,

$$\begin{aligned}
\mathcal{L}(l) &:= \frac{p(l \mid \overline{y}, n, m, s, N)}{p(l \mid \underline{y}, n, m, s, N)} \\
&= \frac{B(l + n\overline{y} + s, m - l + n(1 - \overline{y}) + N - s)B(n\underline{y} + s, n(1 - \underline{y}) + N - s)}{B(n\overline{y} + s, n(1 - \overline{y}) + N - s)B(l + n\underline{y} + s, m - l + n(1 - \underline{y}) + N - s)} \\
&= \frac{\Gamma(l + n\overline{y} + s)\Gamma(m - l + n(1 - \overline{y}) + N - s)\Gamma(n\underline{y} + s)\Gamma(n(1 - \underline{y}) + N - s)}{\Gamma(l + n\underline{y} + s)\Gamma(m - l + n(1 - \underline{y}) + N - s)\Gamma(n\overline{y} + s)\Gamma(n(1 - \overline{y}) + N - s)} \\
&= \begin{cases}
\dfrac{\prod_{x=0}^{m-1}(x + n(1 - \overline{y}) + N - s)}{\prod_{x=0}^{m-1}(x + n(1 - \underline{y}) + N - s)} & \text{for } l = 0 \\
\dfrac{\prod_{x=0}^{l-1}(x + n\overline{y} + s)\prod_{x=0}^{m-l-1}(x + n(1 - \overline{y}) + N - s)}{\prod_{x=0}^{l-1}(x + n\underline{y} + s)\prod_{x=0}^{m-l-1}(x + n(1 - \underline{y}) + N - s)} & \text{for } 0 < l < m \\
\dfrac{\prod_{x=0}^{m-1}(x + n\overline{y} + s)}{\prod_{x=0}^{m-1}(x + n\underline{y} + s)} & \text{for } l = m
\end{cases}
\end{aligned}$$

since $\Gamma(x + 1) = x\Gamma(x)$.

Thus,

$$\begin{aligned}
\frac{\mathcal{L}(l + 1)}{\mathcal{L}(l)} &= \frac{(l + n\overline{y} + s)(m - l - 1 + n(1 - \underline{y}) + N - s)}{(l + n\underline{y} + s)(m - l - 1 + n(1 - \overline{y}) + N - s)} \\
&> 1 \quad \text{when} \quad 0 \leq \underline{y} < \overline{y} \leq 1
\end{aligned}$$

Hence, $\mathcal{L}(\cdot)$ is monotone increasing for $0 < \underline{y} < \overline{y} < 1$, so that $\beta_{\overline{y}}$ is larger than or equal to $\beta_{\underline{y}}$ in monotone likelihood ratio order ($\beta_{\overline{y}} \geq_{\text{lr}} \beta_{\underline{y}}$). But, $\beta_{\overline{y}} \geq_{\text{lr}} \beta_{\underline{y}} \implies \beta_{\overline{y}} \geq_{\text{st}} \beta_{\underline{y}}$ ([23, Theorem 1.C.1, p.43]) giving the required result. □

Proof of Theorem 2, p16. Consider the likelihood ratio for the two Beta Binomial distributions $\beta_{\overline{n}}$ and $\beta_{\underline{n}}$,

$$
\begin{aligned}
\mathcal{L}(l) &:= \frac{p(l \mid y, \overline{n}, m, s, N)}{p(l \mid y, \underline{n}, m, s, N)} \\
&= \frac{B(l + \overline{n}y + s, m - l + \overline{n}(1-y) + N - s)B(\underline{n}y + s, \underline{n}(1-y) + N - s)}{B(\overline{n}y + s, \overline{n}(1-y) + N - s)B(l + \underline{n}y + s, m - l + \underline{n}(1-y) + N - s)} \\
&= \frac{\Gamma(l + \overline{n}y + s)\Gamma(m - l + \overline{n}(1-y) + N - s)}{\Gamma(l + \underline{n}y + s)\Gamma(m - l + \underline{n}(1-y) + N - s)} \\
&\quad \times \frac{\Gamma(\underline{n}y + s)\Gamma(\underline{n}(1-y) + N - s)\Gamma(\overline{n} + N)\Gamma(m + \underline{n} + N)}{\Gamma(\overline{n}y + s)\Gamma(\overline{n}(1-y) + N - s)\Gamma(\underline{n} + N)\Gamma(m + \overline{n} + N)} \\
&= \begin{cases}
\dfrac{\prod_{x=0}^{m-1}(x + \overline{n}(1-y) + N - s)\prod_{x=0}^{m-1}(x + \underline{n} + N)}{\prod_{x=0}^{m-1}(x + \underline{n}(1-y) + N - s)\prod_{x=0}^{m-1}(x + \overline{n} + N)} & \text{for } l = 0 \\
\dfrac{\prod_{x=0}^{l-1}(x + \overline{n}y + s)\prod_{x=0}^{m-l-1}(x + \overline{n}(1-y) + N - s)}{\prod_{x=0}^{l-1}(x + \underline{n}y + s)\prod_{x=0}^{m-l-1}(x + \underline{n}(1-y) + N - s)} & \\
\qquad \times \dfrac{\prod_{x=0}^{m-1}(x + \underline{n} + N)}{\prod_{x=0}^{m-1}(x + \overline{n} + N)} & \text{for } 0 < l < m \\
\dfrac{\prod_{x=0}^{m-1}(x + \overline{n}y + s)\prod_{x=0}^{m-1}(x + \underline{n} + N)}{\prod_{x=0}^{m-1}(x + \underline{n}y + s)\prod_{x=0}^{m-1}(x + \overline{n} + N)} & \text{for } l = m
\end{cases}
\end{aligned}
$$

since $\Gamma(x+1) = x\Gamma(x)$.

Thus,

$$\frac{\mathcal{L}(l+1)}{\mathcal{L}(l)} = \frac{(l + \overline{n}y + s)(m - l - 1 + \underline{n}(1-y) + N - s)}{(l + \underline{n}y + s)(m - l - 1 + \overline{n}(1-y) + N - s)}$$

However, unlike the case for the y parameter in Theorem 1, neither $\beta_{\underline{n}}$ nor $\beta_{\overline{n}}$ can be guaranteed to dominate for all possible values for the other parameters, so that necessary conditions for monotonicity (either increasing or decreasing) must be established. We require,

$$\frac{(l + \overline{n}y + s)(m - l - 1 + \underline{n}(1-y) + N - s)}{(l + \underline{n}y + s)(m - l - 1 + \overline{n}(1-y) + N - s)} > 1$$

After extensive routine algebra, this can be conveniently expressed as

$$(\overline{n} - \underline{n})[y(N + m - 1) - s] - l(\overline{n} - \underline{n}) > 0.$$

This limit is hardest to satisfy for $l = m - 1$ since $\overline{n} - \underline{n} > 0$ (note $l \neq m$ since we are evaluating $\mathcal{L}(l+1)/\mathcal{L}(l)$, so $m-1$ is the maximal value l can take). Thus, for monotonicity to hold for all l, we require

$$\begin{aligned}(\overline{n} - \underline{n})[y(N+m-1) - s] - (m-1)(\overline{n} - \underline{n}) > 0 \\ \implies (\overline{n} - \underline{n})[y(N+m-1) - s - m + 1] > 0\end{aligned}$$

Since $\overline{n} - \underline{n} > 0$ by definition, we have a monotonically increasing likelihood ratio only when

$$y(N+m-1) - s - m + 1 > 0.$$

By a similar argument, the likelihood ratio is only monotonically decreasing when

$$y(N+m-1) - s < 0.$$

Thus,

$$y > \frac{s+m-1}{N+m-1} \implies \beta_{\overline{n}} \geq_{\text{lr}} \beta_{\underline{n}} \implies \beta_{\overline{n}} \geq_{\text{st}} \beta_{\underline{n}} \tag{A.1}$$

and

$$y < \frac{s}{N+m-1} \implies \beta_{\overline{n}} \leq_{\text{lr}} \beta_{\underline{n}} \implies \beta_{\overline{n}} \leq_{\text{st}} \beta_{\underline{n}} \tag{A.2}$$

by [23, Theorem 1.C.1, p.43]. In the intermediate case,

$$\frac{s}{N+m-1} < y < \frac{s+m-1}{N+m-1}$$

standard likelihood ratio ordering theory cannot definitively state the stochastic ordering on $\beta_{\overline{n}}$ and $\beta_{\underline{n}}$. □

Proof of Lemma 3, p17. (A.1) and (A.2) are sufficient but not necessary conditions. Using theory in [19] we can sharpen these conditions to provide first-order stochastic dominance conditions for a larger range of parameter values.

Proposition 2.1, p.399 of [19] proves that half-monotone likelihood ratio ordering — i.e. monotonicity of $\mathcal{L}(l+1)/\mathcal{L}(l)$ — together with left and right tail conditions on $\mathcal{L}(\cdot)$ imply first order stochastic dominance.

Half-monotonicity of $\mathcal{L}(\cdot)$

Although there exist parameters for which $\mathcal{L}(\cdot)$ is not monotone, it is half-monotone. That is, $\mathcal{L}(l+1)/\mathcal{L}(l)$ is itself monotone. For simplicity, write

$$\frac{\mathcal{L}(l+1)}{\mathcal{L}(l)} = \frac{(l+\overline{\psi})(\underline{\eta} - l)}{(l+\underline{\psi})(\overline{\eta} - l)} \quad \text{where} \begin{cases} \overline{\psi} = \overline{n}y + s \\ \underline{\psi} = \underline{n}y + s \\ \overline{\eta} = m - 1 + \overline{n}(1-y) + N - s \\ \underline{\eta} = m - 1 + \underline{n}(1-y) + N - s \end{cases}$$

Then,

$$\frac{\mathcal{L}(l+2)/\mathcal{L}(l+1)}{\mathcal{L}(l+1)/\mathcal{L}(l)} = \frac{(\underline{\psi}+l)(\overline{\psi}+l+1)(\underline{\eta}-l-1)(\overline{\eta}-l)}{(\underline{\psi}+l+1)(\overline{\psi}+l)(\underline{\eta}-l)(\overline{\eta}-l-1)} < 1$$

$$\iff \qquad \frac{\underline{\psi}+l}{\overline{\psi}+l} \cdot \frac{\underline{\eta}-l-1}{\overline{\eta}-l-1} < \frac{\underline{\psi}+l+1}{\overline{\psi}+l+1} \cdot \frac{\underline{\eta}-l}{\overline{\eta}-l}$$

But, $\overline{\psi} > \underline{\psi} > 0$, $\overline{\eta} > \underline{\eta} > 0$, $l > 0$, so it is trivial to prove

$$\frac{\underline{\psi}+l}{\overline{\psi}+l} < \frac{\underline{\psi}+l+1}{\overline{\psi}+l+1} \qquad \text{and} \qquad \frac{\underline{\eta}-l-1}{\overline{\eta}-l-1} < \frac{\underline{\eta}-l}{\overline{\eta}-l} \qquad \forall\, l \in \{0,\ldots,m\}$$

Thus we can conclude that $\mathcal{L}(\cdot)$ is half monotone decreasing.

Tail conditions on $\mathcal{L}(\cdot)$

It is not difficult to derive the same loose bounds as in Theorem 2 using the tail conditions. However, it is also easy to see that these are sufficient but not necessary. Sharpening these bounds in terms of the other parameter values involves seemingly intractable algebra, so we leave the tail condition as the alternative slightly more costly numerical check when the conditions of Theorem 2 are not satisfied. Evaluation of $\mathcal{L}(\cdot)$ at two values is still orders of magnitude less costly than reevaluation of $\underline{R}_{\text{sys}}(\cdot)$ or $\overline{R}_{\text{sys}}(\cdot)$. □

B. Theory versus optimisation

The above theory substantially reduces the amount that numerical optimisation must be used. Herein, we examine more carefully precisely how often the theory cannot be applied, resulting in numerical optimisation being used.

Given some unknown functioning probability p, which gives rise to the observed data s, the probability that Theorem 2 does not apply is

$$P(y(N+m-1)-m+1 < s < y(N+m-1)) = \sum_{i=\lceil y(N+m-1)-m+1 \rceil}^{\lfloor y(N+m-1) \rfloor} \binom{N}{i} p^i (1-p)^{N-i}$$

Likewise the probability that Lemma 3 does not apply is

$$P(\mathcal{L}_{\overline{n},\underline{n}}(0) \lesseqgtr 1 \cap \mathcal{L}_{\overline{n},\underline{n}}(m) \gtreqless 1) = \sum_{s \in \mathcal{S}} \binom{N}{s} p^s (1-p)^{N-s}$$

$$\text{where} \quad \mathcal{S} = \left\{ s : \frac{p(0 \mid y, \overline{n}, m, s, N)}{p(0 \mid y, \underline{n}, m, s, N)} \lesseqgtr 1 \cap \frac{p(m \mid y, \overline{n}, m, s, N)}{p(m \mid y, \underline{n}, m, s, N)} \gtreqless 1 \right\}$$

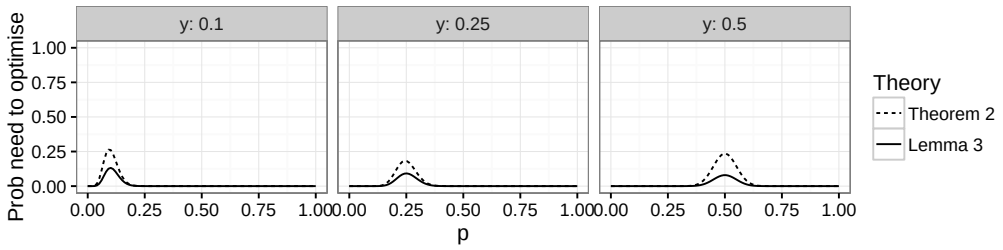

Figure B.9: Plots of the probability that numerical optimisation is required when $N = 100$, $m = 3$, $\underline{n} = 1$ and $\overline{n} = 5$ for changing choices of prior, y, as a function of the unknown functioning probability, p.

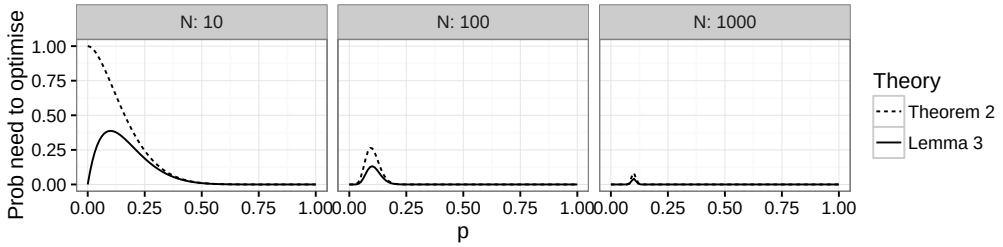

Figure B.10: Plots of the probability that numerical optimisation is required when $y = 0.1$, $m = 3$, $\underline{n} = 1$ and $\overline{n} = 5$ for changing amounts of test data, N, as a function of the unknown functioning probability, p.

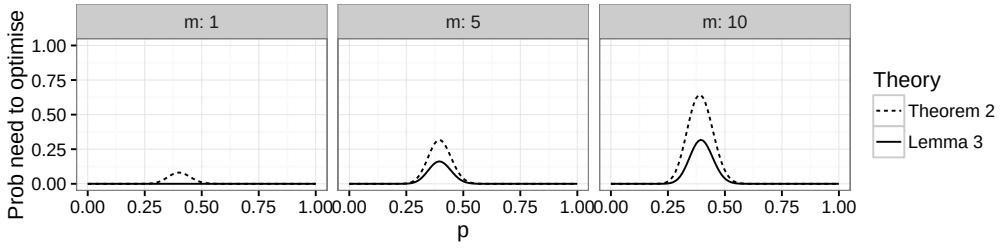

Figure B.11: Plots of the probability that numerical optimisation is required when $y = 0.4$, $N = 100$, $\underline{n} = 1$ and $\overline{n} = 5$ for differing numbers of components in the system, m, as a function of the unknown functioning probability, p.

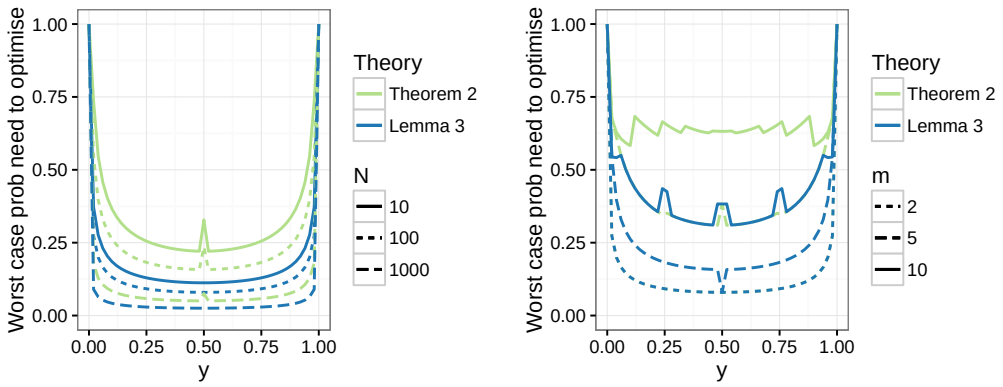

Figure B.12: Plots of the worst case probability that numerical optimisation is required for different choices of y, when $m = 3$ (left), $N = 100$ (right), $\underline{n} = 1$ and $\overline{n} = 5$. There are some artefacts due to the underlying discrete Binomial probabilities.

B.1. Some example scenarios

With the prior value y (upper or lower) fixed, the above expressions depend only on the unknown functioning probability p, which gives rise under repeated experiments to an observed number of functioning components, s.

Thus, Figures B.9, B.10 and B.11 show how often optimisation would be required under long term repeated use of this technique.

The worst case scenario is that the prior, y, is exactly equal to the true unknown probability of functioning, p. Regardless of y, optimisation is required less frequently with larger test dataset sizes (increasing N), but offset by increasing numbers of components of the same type in a system (increasing m). Not shown is that $\underline{n}$ and $\overline{n}$ clearly do not affect the probability that Theorem 2 does not apply, but also the impact on Lemma 3 is negligible.

B.2. Worst case scenario

Clearly the true functioning probability, p, is unknown. Figure B.12 therefore shows the worst case scenario for any choice of y — that is, where the maximum of each curve from the previous subsection is taken.

Especially given the upper and lower choices of y, it may be reasonable to suppose that it is unlikely that the prior probability will all that often exactly equal the true failure probability (at least not for long stretches of time). Therefore, a final pair of plots replicates the analysis in Figure B.12, but where the worst case p is subject to the constraint $|y - p| > 0.1$. This is shown in Figure B.13

Note that all these analyses are for evaluation of a single fixed time point at which $\underline{R}_{\text{sys}}(\cdot)$ or $\overline{R}_{\text{sys}}(\cdot)$ is computed. Of relevance to the overall compute

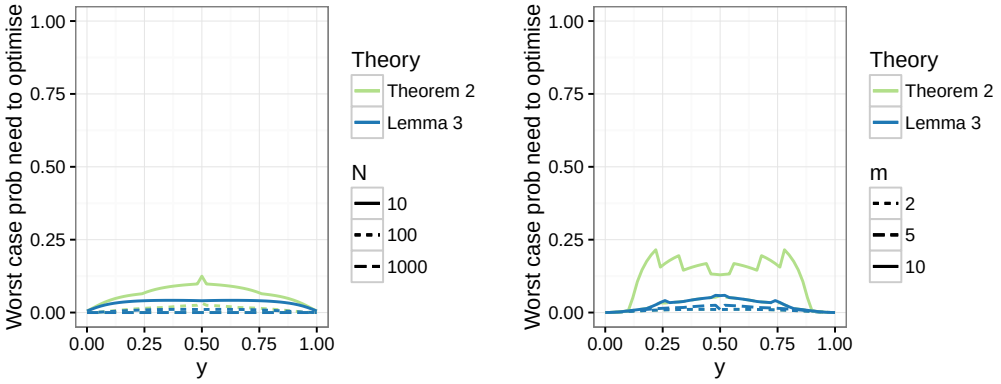

Figure B.13: Plots of the worst case probability that numerical optimisation is required for different choices of y, when $m = 3$ (left), $N = 100$ (right), $\underline{n} = 1$ and $\overline{n} = 5$, subject to the constraint $|y - p| > 0.1$. There are some artefacts due to the underlying discrete Binomial probabilities.

time is the behaviour over time, and it seems unlikely that the worst case scenario will occur for long runs of the times under analysis.

B.3. Theoretical analysis for brake system example

Such temporal aspects are now considered by analysing the brake system example from Section 7.2.2. As noted in Section 7.2.2, the grid contains 301 time points and the system has 4 types of components, meaning stochastic dominance must be verified via theory or optimisation $1,204$ times for each of $\underline{R}_{\text{sys}}(\cdot)$ and $\overline{R}_{\text{sys}}(\cdot)$.

Figure B.14 shows the worst case probabilities (subject to $|y - p| > 0.1$) as they change over time in the brake system example. Displayed are the worst case probabilities that optimisation will be required for the choice of prior and the values of m and N as in Section 7.2.2, but without using the test data (which define s). Note in particular that when $m = 1$ (as is the case for component types M and H), Lemma 3 always applies for $y \in (0, 1)$ and so optimisation is never required. This is true for all values of N and y.

For the test data as used in Section 7.2.2, theory could not determine stochastic dominance in 98 (or 8.1%) of cases for $\underline{R}_{\text{sys}}(\cdot)$, while it was in 52 (or 4.3%) of cases for $\overline{R}_{\text{sys}}(\cdot)$.

C. Software details

Functions which make it easy to use the methods of this paper have been added to the R package `ReliabilityTheory` [2]. There are two functions of particular note: `computeSystemSurvivalSignature` and, implementing the result from Appendix A above, `nonParBayesSystemInferencePriorSets`.

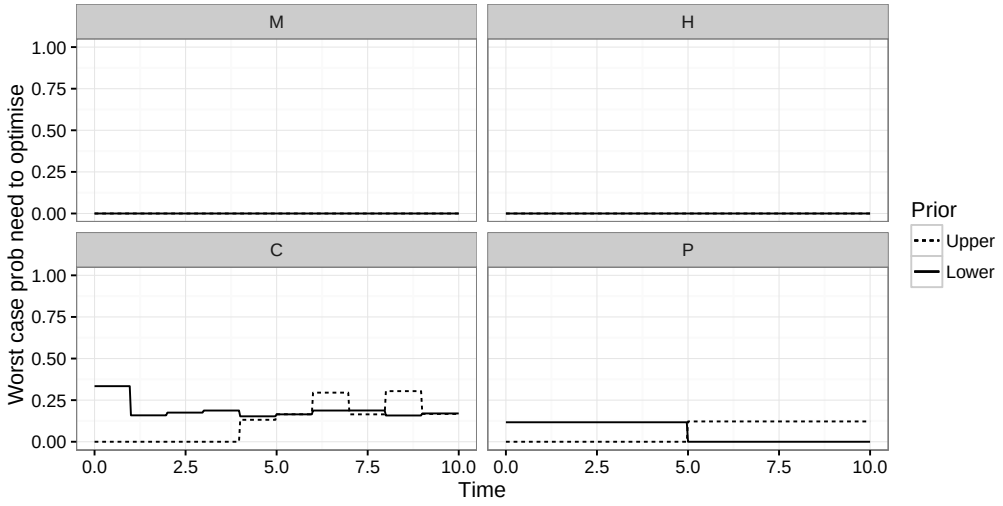

Figure B.14: Plots of the worst case probability that numerical optimisation is required over time in the brake example, subject to the constraint $|y - p| > 0.1$.

C.1. Computing the survival signature

The function `computeSystemSurvivalSignature` allows easy computation of the survival signature if the system is expressed as an undirected graph with 'start' and 'terminal' nodes (which are not considered components for survival signature computation). The system is considered to work if there is a path from the start to the terminal node passing only through functioning components.

Graph representations of systems are most simply defined by using the `graph.formula` function. The 'start' node should be denoted `s` and the 'terminal' node should be denoted `t` and intermediate nodes (representing actual components) should be numbered and connected by edges denoted by `-`, where the numbering denotes physically distinct components. Component numbers can be repeated to include multiple links. For example, to build a simple three component series system:

```
sys <- graph.formula(s-1-2-3-t)
```

and to build a three component parallel system:

```
sys <- graph.formula(s-1-t, s-2-t, s-3-t)
```

There is an additional shorthand which indicates a link exists to a list of multiple components separated by the `:` operator, so that the parallel system can be also be expressed more compactly by:

```
sys <- graph.formula(s-1:2:3-t)
```

Therefore, the simple bridge system of Figure 4 can be constructed with:

```
sys <- graph.formula(s-1-2-3-t, s-4-5-3-t, 1:4-6-2:5)
```

- `s-1-2-3-t` signifies the route from left to right entering the first component going across the top of the system block diagram in Figure 4;
- `s-4-5-3-t` signifies the bottom route through the block diagram;
- `1:4-6-2:5` connects the top two components of type 3 to the bottom two components of type 3, signifying the bridge.

Naturally such as expression is not necessarily unique, so that completely equivalently one may write:

```
sys <- graph.formula(s-1:4-6-2:5-3-t, 1-2, 4-5)
```

With the structure defined and the individual components numbered, it just remains to specify the types of each component. This can be done using the `setCompTypes` function. This function takes the system graph and a list of component type names (as the tag) and corresponding component numbers (as the value). Thus, completing the example for Figure 4:

```
sys <- setCompTypes(sys, list("T1"=c(1,2,4,5), "T2"=c(6),
                              "T3"=c(3)))
```

Computing the survival signature then involves a simple function call:

```
survsig <- computeSystemSurvivalSignature(sys)
```

C.2. Computing sets of system survival probabilities

Once the system has been correctly described using an undirected graph as above, the methods presented in Sections 3 – 6 can be used via the function `nonParBayesSystemInferencePriorSets`.

The function prototype is:

```
nonParBayesSystemInferencePriorSets(at.times, survival.signature,
          test.data, nLower=2, nUpper=2, yLower=0.5, yUpper=0.5)
```

Aside from the system design, which can be passed to the function via the `survival.signature` argument, the remaining elements which must be specified are the:

1. grid of times at which to evaluate the posterior, $\mathcal{T} = \{t_1, \ldots, t_{\max}\}$, via the `at.times` argument.
2. component test data $\boldsymbol{t}^k = (t^k_1, \ldots, t^k_{n_k})$ for $k = 1, \ldots, K$, via the `test.data` argument.
3. prior sets via the range of prior parameter sets $\Pi^{(0)}_{k,t} = [\underline{n}^{(0)}_{k,t}, \overline{n}^{(0)}_{k,t}] \times [\underline{y}^{(0)}_{k,t}, \overline{y}^{(0)}_{k,t}]$, via the `nLower`, `nUpper`, `yLower` and `yUpper` arguments.

The grid of times, `at.times`, is specified as simply a vector of time points.

The `test.data` argument is a list of component type names (as the tag) and corresponding lifetime data (as the value), for example a toy sized dataset for each component would be expressed as:

```
test.data=list("T1"=c(0.19, 0.73, 1.87, 1.17),
               "T2"=c(0.22, 0.27, 0.63, 1.80, 1.25, 1.95),
               "T3"=c(1.33, 0.65, 1.59))
```

Finally, there are multiple options for specifying the prior parameter sets. Each of the `nLower`, `nUpper`, `yLower` and `yUpper` arguments can be specified as:

- a single value for a homogeneous prior across time and components. e.g. `nLower=2` $\implies \underline{n}_{k,t}^{(0)} = 2 \; \forall \, k, t$
- a vector of values of length $|\mathcal{T}|$ (`length(at.times)`), for a time inhomogeneous prior which is identical across component types.
- a data frame of size $1 \times K$, where each column is named the same as in the `survival.signature` and `test.data` arguments, for a time homogeneous prior which varies across component types.
- a data frame of size $|\mathcal{T}| \times K$, where each column is named the same as in the `survival.signature` and `test.data` arguments, for a time inhomogeneous prior which varies across component types.

With these arguments supplied, `nonParBayesSystemInferencePriorSets` will then compute the posterior sets automatically in parallel across the cores of a multicore CPU and return a list with two objects, named `lower` and `upper`, containing respectively the lower and upper bound for the system reliability function $R_{\text{sys}}(t)$.